IFAMR

The Value of Big Data in Agriculture Inputs, Farming, and Processing

Special Issue

International Food and Agribusiness Management Review

The IFAMR (ISSN #: 1559-2448) is published quarterly and is available at http://www.ifama.org. For copyright and publishing information, please contact: Kathryn White, Administrative Editor • IFAMA Business Office • 5775 Wayzata Blvd. Suite 700, Minneapolis MN 55416 USA • Tel: 1 (763) 412-1988 • Fax: 1 (763) 971-7958 • Email: ifamr@ifama.org • Web: www.ifama.org

The International Food and Agribusiness Management Review

Volume 19 Special Issue A

The Value of Big Data in Agriculture: Inputs, Farming and Processing

Special Issue Editors

Lead Editor: Eric Jackson, Conservis
Co-Editors: Steven Sonka, University of Illinois
Michael Boehlje, Purdue University
Charles Linville, Ploughman Analytics
Kenneth S. Zuckerberg, Rabobank

International Food and Agribusiness Management Review
Special Issue - Volume 19 Issue A, 2016

TABLE OF CONTENTS

International Food and Agribusiness Management Review
Special Issue - Volume 19 Issue A, 2016

The Value of Big Data in Agriculture: Inputs, Farming and Processing

EDITOR INTRODUCTION

Eric Jackson[ⓘ]

Cofounder and Chief Sustainability Officer, Conservis Corp, Minneapolis, MN, USA

The promise of *big data in agriculture* is very alluring. After all, agriculture is one of the last great enterprises on the planet that hasn't been fully digitized and analyzed. It is a biological manufacturing system, wrought with all the complexities one might expect from jamming humans, machines, natural systems, chemistry, biology, weather and climate into a single box. As Donald Rumsfeld famously quipped "As we know, there are known knowns; there are things we know we know. We also know there are known unknowns; that is to say we know there are some things we do not know. But there are also unknown unknowns—the ones we don't know we don't know."

As I travel the world and discuss the opportunities and challenges of big data in agriculture with other global agriculturalists, several recurring themes are becoming prevalent.

First, agriculture is a very location-specific enterprise. Soil, water and land characteristics—arguably three of the strongest determinates of outcomes—are hyper-local in their variability. No two fields or paddocks or plots are exactly the same.

Second, weather and climate are highly localized. No two growing seasons are the same and the local variability within a season can be very stark.

Third, the proximity of a given farming operation to the marketplace and the transportation infrastructure which enables the handling, movement and storage of crops varies dramatically from location to location.

[ⓘ]Corresponding author: Tel: + 1 612.424.6300
Email: E. Jackson: ejackson@conserviscorp.com

And last but not least, farming methods and practices are as individualized as the humans performing them. There are deep, local, and cultural roots that can drive behavior and decisions made on the farm.

So as we discuss the value of *big data in agriculture*, one has to wonder if small data, i.e. local field-specific data, isn't the key that might unlock the value in the big data vault. Put another way, there are things that we do know that might help uncover that which we don't know. And it's very important that we use all the data tools at our disposal to address the core challenge; the Food and Agricultural Organization of the United Nations has forecasted the need for a 70% increase in global food production by mid-century.

This Special Edition has two primary sections. The first two papers were invited from two of the co-editors on this project. Dr. Steven Sonka starts by framing the characteristics of big data. Then, Dr. Michael Boehlje offers perspectives on how big data might impact industry structure and enhance business margins, particularly in developed agricultural economies.

The next section contains ten peer-reviewed submissions with topics ranging from cattle production; to data privacy; wireless broadband; and food safety, with authors spanning every continent from India, Africa, North and South America and Europe. Each offers us a birds-eye view of big data from both developed and developing economies.

But even with these contributions, and the 20+ other papers that were reviewed in the process of putting this issue together, the fact remains that we can only imagine more than we can know about the value of big data in agriculture. It will be an exciting journey for those of us who choose to climb onboard!

**A special thanks to our co-editors: Steven Sonka, Michael Boehlje, Charlie Linville and Kenneth Zuckerberg for their contributions in helping to bring this issue into fruition.*

International Food and Agribusiness Management Review
Special Issue - Volume 19 Issue A, 2016

Big Data Characteristics

Steven Sonka[①]

Emeritus Chaired Professor of Agricultural Strategy, University of Illinois, 326 Mumford Hall, 1301 W. Gregory Drive, Urbana, IL 61801, USA

Often, *big data* is referred to as a singular entity. It is not! In reality, big data is much more a capability than a thing. It is the capability to extract information and insights where previously it was economically, if not technically, possible to do so. Advances across several technologies are fueling the growing big data capability. These include, but are not limited to computation, data storage, communications, and sensing. The growing ability of analysts and managers to exploit the information provided by the big data capability is equally important.

Only recently have numerous attempts been made to define big data. For example:

- The phrase "big data" refers to large, diverse, complex, longitudinal, and/or distributed data sets generated from instruments, sensors, Internet transactions, e-mail, video, click streams, and/or all other digital sources available today and in the future (The National Science Foundation 2012).

- Big data shall mean the datasets that could not be perceived, acquired, managed and processed by traditional IT and software/hardware tools within a tolerable time (Chen et al. 2014).

- Big data is where the data volume, acquisition velocity, or data representation [variety] limits the ability to perform effective analysis using traditional relational approaches or requires the use of significant horizontal scaling for efficient processing (Cooper and Mell 2012).

- Big data is high-volume, -velocity, and -variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making (Gartner IT Glossary 2012).

[①]Corresponding author: Tel: + 1 217.333.1810
Email: S.S. Sonka: ssonka@illinois.edu

The goal of this paper is to move beyond those definitions to explore the characteristics of big data which have particular relevance in fostering the creation of value in agriculture.

Dimensions of Big Data

Three dimensions (Figure 1) often are employed to describe the big data phenomenon: *volume, velocity, and variety* (Manyika et al. 2011). Each dimension presents both challenges for data management and opportunities to advance agri-business decision-making. These three dimensions focus on the nature of data. However, just having data isn't sufficient. Analytics is the hidden, "secret sauce" of big data. Analytics (discussed later), refers to the increasingly sophisticated means by which useful insights can be fashioned from available data.

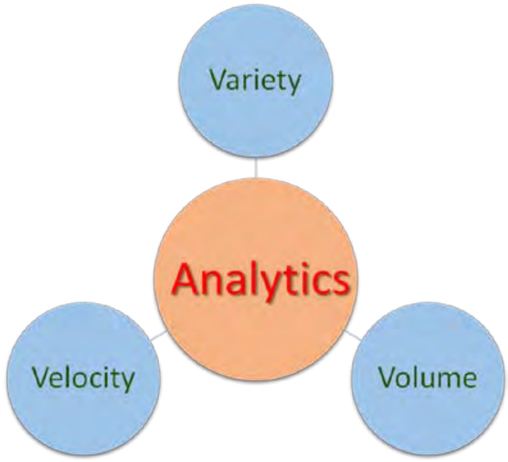


Figure 1. Dimensions of Big Data

Volume: According to IBM (2012) 90% of the data in the world today has been created in the last two years alone. In recent years, statements similar to IBM's observation and its emphasis on volume of data have become increasingly more common.

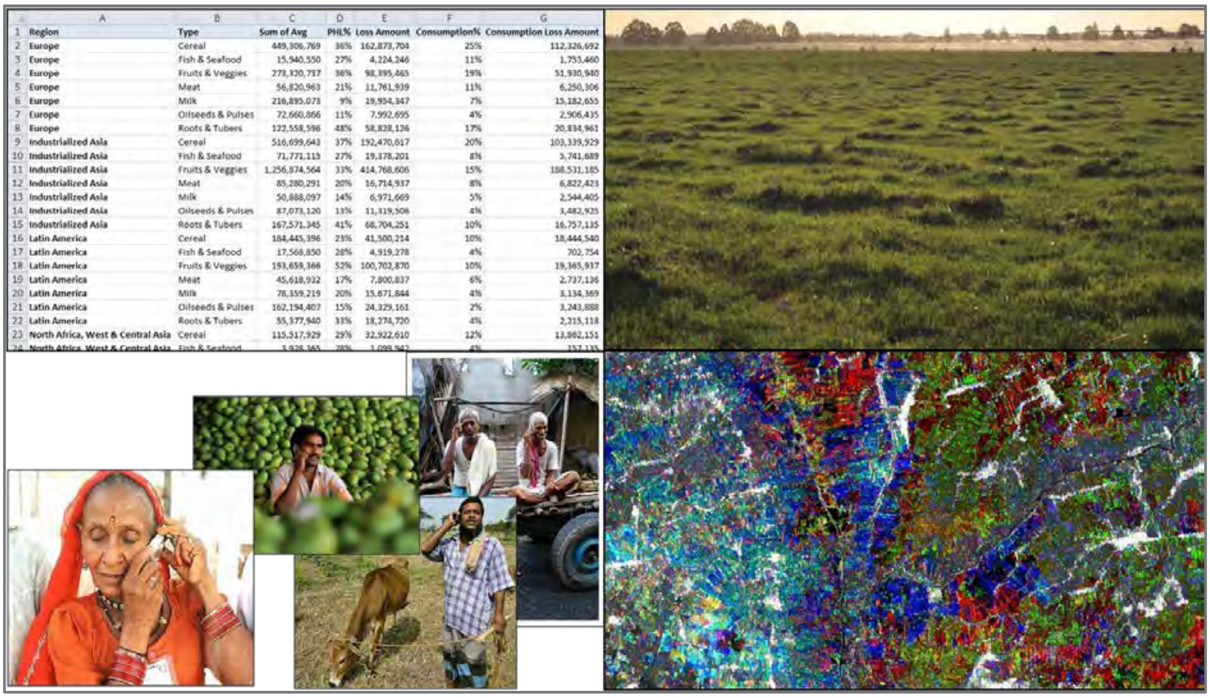
The volume dimension of big data is not defined in specific quantitative terms. Rather, big data refers to datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze. This definition is intentionally subjective; with no single standard of how big a dataset needs to be to be considered big—and that standard can vary between industries and applications.

Velocity: The velocity dimension refers to the capability of understanding and responding to events *as they occur*. Sometimes it's not enough just to know what's happened; rather we want to know what's happening. For example, applications like Google Maps provide real-time traffic information at our fingertips. Google Maps provides live traffic information by analyzing the speed of phones using the Google Maps app on the road (Barth 2009). Based on the changing traffic status and extensive analysis of factors that affect congestion, Google Maps can suggest alternative routes in real-time to ensure a faster and smoother drive.

Variety: As a dimension of big data, variety may be the most novel and intriguing. For many of us, data refers to numbers meaningfully arranged in rows and columns. For big data, the reality of "what is data" is wildly expanding. For example, the movement of your eyes as you read this text could be captured and employed as data.

Suddenly (at least in agricultural measurement terms), the "what is data" question—the variety dimension of big data—has new answers. Figure 2 provides a visual illustration of the change. In its upper left hand corner, we see data as we are used to it – rows and columns of nicely organized numbers. The picture in the upper right hand corner is of a pasture in New Zealand. Pastures are the primary source of nutrition for dairy cows in that country and supplemental fertilization is a necessary economic practice. The uneven pattern of the forage in that field is measured by a sensor on the fertilizer spreader to regulate how much fertilizer is applied—as the spreader goes across the field. In this situation, uneven forage growth is now data. (This also is an example of velocity where the measurement activity is directly linked to action based upon the measurement.)

The lower left hand corner of Figure 2 shows the most versatile sensor in the world – individuals using their cell phone. Particularly for agriculture in developing nations, the cell phone is a phenomenal source of potential change—because of both the information sent to those individuals and information they now can provide. And as illustrated in the lower right hand quadrant of Figure 2, satellite imagery can measure temporal changes in reflectivity of plants to provide estimates of growth (RIICE 2013). The picture is focused on rice production in Asia.



	A	B	C	D	E	F	G
1	Region	Type	Sum of Avg	PHL%	Loss Amount	Consumption%	Consumption Loss Amount
2	Europe	Cereal	449,306,769	36%	162,873,704	25%	112,326,692
3	Europe	Fish & Seafood	15,940,550	27%	4,224,246	11%	1,753,460
4	Europe	Fruits & Veggies	273,320,737	36%	98,395,465	19%	51,930,940
5	Europe	Meat	56,820,963	21%	11,761,939	11%	6,250,306
6	Europe	Milk	216,895,073	9%	19,954,347	7%	15,182,655
7	Europe	Oilseeds & Pulses	72,660,866	11%	7,992,695	4%	2,906,435
8	Europe	Roots & Tubers	122,558,596	48%	58,828,126	17%	20,834,961
9	Industrialized Asia	Cereal	516,699,643	37%	192,470,617	20%	103,339,929
10	Industrialized Asia	Fish & Seafood	71,771,115	27%	19,378,201	8%	5,741,689
11	Industrialized Asia	Fruits & Veggies	1,256,874,564	33%	414,768,606	15%	188,531,185
12	Industrialized Asia	Meat	85,280,291	20%	16,714,937	8%	6,822,423
13	Industrialized Asia	Milk	50,888,097	14%	6,971,669	5%	2,544,405
14	Industrialized Asia	Oilseeds & Pulses	87,073,120	13%	11,319,506	4%	3,482,925
15	Industrialized Asia	Roots & Tubers	167,571,345	41%	68,704,251	10%	16,757,135
16	Latin America	Cereal	184,445,396	23%	41,500,214	10%	18,444,540
17	Latin America	Fish & Seafood	17,568,850	28%	4,919,278	4%	702,754
18	Latin America	Fruits & Veggies	193,659,366	52%	100,702,870	10%	19,365,937
19	Latin America	Meat	45,618,932	17%	7,800,837	6%	2,737,136
20	Latin America	Milk	78,359,219	20%	15,671,844	4%	3,134,369
21	Latin America	Oilseeds & Pulses	162,194,407	15%	24,329,161	2%	3,243,888
22	Latin America	Roots & Tubers	55,377,940	33%	18,274,720	4%	2,215,118
23	North Africa, West & Central Asia	Cereal	115,517,929	29%	32,922,610	12%	13,862,151
24	North Africa, West & Central Asia	Fish & Seafood	3,928,365	28%	1,099,942	4%	157,135

Figure. 2. A few sources of data (Sonka and Cheng 2015).

While satellite imagery is one source of remotely sensed data, recent years have seen a pronounced increase in the capabilities and interest in Unmanned Aerial Systems (UAS) as a source of data for agriculture. There are numerous ongoing efforts to transform UAS technology originally focused on military purposes to applications supporting production agriculture. Further, the potential for proliferation of mini-satellites suggests that remotely captured information may become increasing cost effective for use by agricultural decision makers.

Analytics: Access to lots of data, generated from diverse sources with minimal lag times, sounds attractive. Managers, however, quickly will ask, what do I do with all this stuff? Without similar advances in analytic capabilities, just acquiring more data is unlikely to have significant impact within agriculture. While volume, velocity, and variety are necessary, *analytics* is what allows for fusion across data sources and for new knowledge to be created.

Analytics and its related, more recent term—data science, are key factors by which big data capabilities can contribute to improved performance in the agricultural sector. The differentiating features of big data analytics are 1) inclusion of unstructured and structured data types in combination with 2) extremely large data sets. Data science refers to the study of the generalizable extraction of knowledge from data (Dhar 2013). Tools based upon data science are being developed for implementation in agribusiness, although these efforts are in very early stages.

The concept of analytics is maturing and its uses refined (Davenport 2013; Watson 2013). Analytic efforts can be categorized into one of three types:

- Descriptive efforts focused on documenting what has occurred;
- Predictive efforts exploring what will occur, and;
- Prescriptive efforts identifying what should occur (given the optimization algorithms employed).

In agriculture, as in most fields, descriptive efforts have been most common and even those are relatively infrequent. Within production agriculture, knowing what has occurred—even if very accurately and precisely—does not necessarily provide useful insights as to what should be done in the future.

Production agriculture is complex, where biology, weather, and human actions interact. Science-based methods have been employed to discern why crop and livestock production occurs in the manner in which they do. Indeed, relative to the big data topic, it might be useful to consider these methods as the *small data* process.

The process starts with lab research employing the scientific method as a systematic process to gain knowledge through experimentation. Indeed the scientific method is designed to ensure that the results of an experimental study did not occur just by chance (Herren 2014). However, results left in the lab don't lead to innovation and progress in the farm field. In the United States, the USDA, Land Grant universities, and the private sector have collaborated to exploit scientific advances. A highly effective, but distributed, system emerged where knowledge gained in the laboratory was tested and refined on experimental plots and then extended to agricultural producers.

In agriculture, therefore, knowledge from science will need to be effectively integrated within efforts to accomplish the goals of predictive and prescriptive analytics. Even with this additional complication, the potential of tools based upon emerging data science capabilities offers significant promise to more effectively optimize operations and create value within the agricultural sector.

References

Barth, D. 2009. The bright side of sitting in traffic: Crowdsourcing road congestion data. Google's Official Blog: https://googleblog.blogspot.com/2009/08/bright-side-of-sitting-in-traffic.html

Chen, M., Mao, S., and Y. Liu. 2014. Big data: A survey. *Mobile Networks and Applications* 19(2): 171–209.

Cooper, M., and P. Mell. 2012. Tackling Big Data. National Institute of Standards and Technology. http://csrc.nist.gov/groups/SMA/forum/documents/june2012presentations/fcsm_june2012_cooper_mell.pdf [accessed February 2, 2016].

Davenport, T. H. 2013. Analytics 3.0. *Harvard Business Review* 91(12): 64–72.

Dhar, V. 2013. Data Science and Prediction. *Communications of the ACM* 56 (12): 64–73.

Gartner IT Glossary. 2012. *Big Data*. http://www.gartner.com/it-glossary/big-data.

Herren, R. V. 2014. *The Science of Agriculture A Biological Approach 4th Ed*. Independence: Cengage Learning.

IBM. 2012. *What is big data?* https://www-01.ibm.com/software/data/bigdata/what-is-big-data.html.

Manyika, J., M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, and A. Byers. 2011. *Big Data: The next frontier for innovation, competition, and productivity*. McKinsey & Company website: http://www.mckinsey.com/insights/business_technology /big_data_the_next_frontier_for_innovation. [Accessed February 2, 2016].

National Science Foundation. 2012. Core techniques and technologies for advancing big data science & engineering. National Academies Press. http://www.nsf.gov/pubs/ 2012/nsf12499/nsf12499.htm. [Accessed February 2, 2016].

Remote sensing-based information and insurance for crops in emerging economics [RIICE]. 2013. About RIICE. International Rice Research Institute. http://www. riice.org/ about-riice/ [Accessed February 2, 2016].

Sonka, S., and Y.-T. Cheng. 2015. Big Data: More Than a Lot of Numbers! *farmdoc daily* Department of Agricultural and Consumer Economics, University of Illinois at Urbana-Champaign. (5):201 http://farmdocdaily.illinois.edu/2015/10/big-data-more-than-a-lot-of-numbers.html.

Watson, H. J. 2013. The Business Case for Analytics. *BizEd* 12 (3): 49–54.

International Food and Agribusiness Management Review
Special Issue - Volume 19 Issue A, 2016

How Might Big Data Impact Industry Structure and Enhance Margins?

Michael Boehlje[①]

Distinguished Professor of Agricultural Economics
Purdue University, Krannert Room 660, West Lafayette, IN 47907-2056 USA

How might *big data* impact the agricultural sector and food industry? The impacts on the structure of the industry and the profit margins of individual businesses are numerous, but two critical impacts are: 1) improvements in supply chain linkages to enhance efficiency and effectiveness of the food production and distribution industry; and 2) improvements in on-farm production practices. This commentary provides a brief synopsis of these two impacts.

Supply Chain Linkages

Consumers, particularly those in the developed economics, are becoming increasingly demanding in terms of the attributes and characteristics of the products they consume. Traditional attributes of plant and animal protein products such as nutritional content, taste, texture, affordability, and safety are still mainstays of consumer's expectations, but their expectations of predictability and reliability have increased. With a specific focus on food safety and quality, it is argued that a whole chain traceability system can reduce exposure to hazardous foods and reduce quality deterioration across the chain from producer to consumer. Big data driven quality/safety/traceability systems provide the capabilities to respond to these increased consumer expectations. Such systems have significant benefits in terms of disease control and management of food contamination as argued by Adam et al. (2016) in this issue.

[①]Corresponding author: Tel: + 1 765.494.4222
Email: M. Boehlje: boehljem@purdue.edu

Other attributes have become more important in shaping consumer buying behavior as well as society's expectations from the food industry—attributes that economists call "credence" attributes are generally harder to measure and often a function of how the product is produced and processed along the entire value chain from breeding/genetics, to retail outlets (traditional grocery stores, restaurants, food service providers, and on-line vendors such as Amazon.com). Such attributes include: additive or antibiotic free, organic production systems, locally and/or family-farmer grown, animal treatment/welfare production practices, sustainable production/processing/distribution systems, etc. Given that many credence attributes are not characteristics of the final product but instead processes and activities that do or do not occur across the value or supply chain, documentation and certification often can only occur through systems of whole-chain tracking and tracing. As a consequence, data and information systems are required that monitor and measure these processes and activities at each stage of the supply chain. Equally important, this data and information must be tagged or linked to the physical product (boxes of cereal, cuts of meat, etc.) that flows along that supply chain so that the final product can be credibly marketed and certified as having the attributes that consumer's desire. Some have argued that the incentives of enhancing food safety, product quality, and traceability to guarantee credence attributes and responsiveness to consumer demands and societal expectations of the food production /processing /distribution system may be more important than production efficiencies at the producer level in incenting adoption of big data technologies/systems in the food industry (Sonka 2016).

But are consumers willing to pay for "credence" attributes that require different and more costly production processes as well as unique and costly (tracking/tracing, segregation, storage and handling, and inventory) management processes along the supply chain from producers to consumers? Numerous studies indicate that at least a segment of meat and animal protein consumers are willing to pay for unique attributes. For example Olynk, Tonsor and Wolf (2010) estimate that consumers would pay a $1.74 per pound premium for pork chops that are USDA – PVP verified that individual crates and stalls are not permitted in the production process. Olynk (2012) also found that consumers are willing to pay for pasture access, non-antibiotic use and non-use of crates and stalls in dairy production. Wolf, Tonsor and Olynk (2011) found that consumers were generally willing to pay substantial premiums for milk produced without the use of rbST, on local family farms, with assured food safety enhancement, when claims are verified by the U.S. Department of Agriculture.

In addition, more systematic alignment along the supply chain from input supplier and manufacturing to food retailer has the potential to increase efficiency through better inventory management and product flow scheduling in both differentiated products and commodity supply chains. This alignment will be facilitated by big data technologies and information systems. For example, the logistics and inventory management challenges across all stages (from grain and livestock production through processing and distribution) have the potential for costly stock-outs as well as excess inventories and (waste/spoilage/ quality) deterioration unless the system is well coordinated. Information and communication systems that facilitate alignment and improve the ability to fulfill current product flows and more accurately predict future shortages, bottlenecks, or excess stock will be increasingly driven by big data analytical programs and systems.

While the verification discussion is primarily relevant in developed countries, extended supply chains are increasingly important in developing country agriculture where urbanization is rapidly redefining how food reaches consumers. Coordination of delivery of inputs to farmers and the collection, distribution, and transformation of agricultural products into food is relatively ineffective and inefficient in the developing compared to developed economies. The phenomenal increase in availability and adoption of cell phones, however, offers a means by which communication and coordination capabilities can be greatly strengthened. Distribution and logistics systems are improving with increased investment in logistics/transportation infrastructure, storage and handling, and cold chain distribution systems. Coupled with big data analytics, systematic improvements in supply chain performance are now potentially available.

On-Farm Production Practices

How might big data technologies/systems enhance the ability of producers of agricultural products to be more precise in their production practices and thus improve efficiency and profitability? This concept of precision farming—using information technology to add exactness to the quantity, quality, timing and location of the application and utilization of inputs in crop and livestock production and to produce specific attribute products/outputs—has been discussed and debated for years. But after more than two decades of innovation in this area, our ability to capitalize on this concept has fallen far short of the potential. For example agricultural retailers in the US estimated in 2015 that 41% of the acres in their market area utilize grid or zone soil sampling procedures. While this is up from 12% in 2000, it's still well below full-adoption levels. Furthermore, agricultural retailers estimated in 2015 that, on average, 32% of acres in their market area utilized variable rate technologies for multiple-nutrient fertilizer applications. While this is up from 3% in 2001, technology adoption has been slow (Erickson and Widmar 2015).

Will big data driven technologies/systems have the ability to cost effectively provide the prescriptions that precision farming requires? Recent advances in measuring / monitoring / sensing technology combined with continued improvement in nutritional and biological technology and process control input application technology make more precise input application and measurement of physical output possible. But do we have adequate precision and accuracy to fulfill the promise? More specifically do we have the scientific and numerical evidence based answers to the following questions?

1. What are the fundamental drivers/determinants/constraints of plant/animal growth and what are the specific structure and parameters of the underlying growth model?

2. What technologies are available to accurately real-time measure/sense/monitor the growth process?

3. How regularly and in real-time can growth conditions, drivers, determinants, and constraints on growth be measured?

4. What are the accuracy and measurement errors in measuring outputs (yield, production) and inputs (seed, nutrition, location/spatial, etc.) in biological growth processes?

5. What are the characteristics of the output distributions (i.e. normal, skewed, etc.)?

6. What are the alternative (application/process) control technologies that can be used in real-time to manage and intervene in order to enhance and control biological growth process?

7. What are the errors/accuracy in “application” technology (seed and fertilizer placement, spray patterns and dosage, tank or batch composition and concentration, etc.)?

8. What data aggregation and sharing is needed to obtain essential insights at the appropriate level of granularity given the long cycle-time in biological manufacturing?

9. What information insights are essential to supply chain partners (buyers and suppliers) to increase producer efficiency and profitability while reducing their risk?

10. How might Bayesian/stochastic/systems dynamics with feedback numerical decision models and “options” modeling concepts that focus on the “tails” of the output distributions be used to assess risks and rewards and obtain insights for improved decisions?

The more accurate and positive the answers we find to these questions, the higher the prospects that big data driven technologies and systems will enhance farmer’s profit margins and thus be more widely adopted.

References

Adam, B. A., R. Holcomb, M. Buser, B. Mayfield, J. Thomas, C. A. O’Bryan, P. Crandall, D. Knipe Richard Knipe and S. C. Ricke. 2016. Enhancing Food Safety, Product Quality and Value-Added in Food Supply Chains using Whole-Chain Traceable. *International Food and Agribusiness Management Review* 19(A):191–213.

Erickson, B. and D. A. Widmar. 2015. 2015 Precision Agricultural Services Dealership Survey Results. West Lafayette, Indiana: Purdue University. Center for Food and Agribusiness, Purdue University. http://agribusiness.purdue.edu/files/resources/2015-crop-life-purdue-precision-dealer-survey.pdf. [accessed June 8, 2016].

Olynk, N.J., 2012. Assessing Changing Consumer Preferences for Livestock Production Practices. *Animal Frontiers* 2(3):32–38.

Olynk, N.J., G. T. Tonsor, and C.A. Wolf. 2010. Consumer Willingness to Pay for Livestock Credence Attribute Claim Verification. *Journal of Agricultural and Resource Economics* 35(2):261–280.

Sonka, Steven. 2016. Big Data: Fueling the Next Evolution of Agricultural Innovation. *Journal of Innovation Management* 4(1):114–136.

Wolf, C. A., G. T. Tonsor, and N. J. Olynk, 2011. Understanding U.S. Consumer Demand for Milk Production Attributes. *Journal of Agricultural and Resource Economics* 36(2):326-342.

International Food and Agribusiness Management Review
Special Issue - Volume 19 Issue A, 2016

A Network-Science Support System for Food Chain Safety: A Case from Hungarian Cattle Production

Ákos Jóźwiak [a], Mátyás Milkovics[b], and Zoltan Lakner[ⓘc]

[a] *DVM, Vice-Director National Food Chain Safety Office, System Management and Supervision Directorate, Keleti Károly u. 24, Budapest, Hungary*

[b] *MSC, National Food Chain Safety Office, System Management and Supervision Directorate, Keleti Károly u. 24, Budapest, Hungary*

[c] *Professor, Szent István University, Faculty of Food Science, Villányi út 35-43, Budapest, Hungary*

Abstract

In a risk analysis framework, food chain safety measures should be objective and scientifically based. Network science – as a decision support tool – may have an important role in bringing safety to the food supply.

The aim of the present work is to develop a network-based assessment methodology for Hungarian cattle holdings. The criteria of which is (1) suitable for risk-based planning in order to put resources into the most critical elements of the cattle production network; (2) should be capable of simulating different epidemiological situations in order to increase preparedness for real epidemics.

Keywords: big data, centrality measures, network analysis, veterinary inspection, risk-based planning

ⓘCorresponding author: Tel: + 36305251896
Email: Z. Lakner: zoltan.lakner@etk.szie.hu
A. Jozwiak: jozwiaka@nebih.gov.hu
M. M. Milkovics: milkovicsm@nebih.gov.hu

Introduction

Cattle breeding and trade have been an important part of economic life in the Carpathian basin since the late Iron Age (Bökönyi 1971). Hungarian landholders were important suppliers of animals to markets in Italy and Poland in the middle ages (Sugar et al. 1994). During the Communist regime, which lasted from after the Second World War to 1989, the overwhelming majority of cattle production was concentrated in large-scale production cooperatives and state farms (Csizmadia 1974). While these state and cooperative farms decreased efficiency and productivity considerably, they made veterinary inspection of bovine herds relatively simple. After the system change in 1989, as a result of the agricultural transition and privatization, the number of bovine herds increased but the professional quality of management remained relatively unchanged (Csáki 1990).

Food chain safety "from farm to fork"—together with its elements such as animal health or food safety—are the focus of both the agri-food industry and the control authorities. With the increasing volume and complexity of international trade, traceability issues have become more important than ever. Additionally, bovine-related veterinary problems (in particular BSE, foot and mouth disease) have increased the importance of veterinary management and inspection of herds worldwide (Nikiforuk 2008). The cattle passport system of the EU, as well as national animal movement detection systems (Dubé et al. 2009), offer the possibility of tracing animal movements.

It is well documented that the herd epidemiology is considerably influenced by the mobility of animals (Kao 2002; Kao et al. 2007). The arrival of new, infected animals on non-infected herds increase the probability of disease transmission. The EU animal health strategy highlights the importance of individual animal identification, supporting legal and financial issues necessary for data collection on animal transportation.

The increasing integration of the Hungarian agri-food system into the EU—as a consequence of both trade liberalization and EU membership—have made the situation even more difficult (Bojnec and Fertő 2009; Coulombier and Takkinen 2013). The food chain safety authority is faced with a mission which is practically impossible to implement using traditional methods: i.e. increasing the effectiveness and reliability of food chain control (veterinary inspection of herds in particular) while at the same time given declining resources (Luning et al. 2015).

The Risk Analysis Framework

It is essential to maintain the health of plants, animals and humans to ensure the chemical and microbiological safety of our food ('food chain safety'), while maintaining the sustainability of agri-food production and trade ('food security'). Food chain industry stakeholders—who have primary responsibility for ensuring safety—need to apply a hazard analysis framework to ensure a process-based, preventative, effective operation. Food chain control authorities, when making decisions on control and intervention issues, must use the risk analysis framework as defined by FAO/WHO (2007).

Risk analysis is used to develop an estimate of the risks to human health and safety (risk assessment); identify and implement appropriate measures to control the risks (risk management); and communicate with stakeholders about the risks and measures applied (risk communication). Risk is defined in this context as a product of the severity of the hazard and the probability of its occurrence. Based on FAO/WHO guidelines, EU member states have to apply a risk-based approach managing food chain safety risks: Article 3 of Regulation 882/2004/EC (European Parliament and the Council 2004) states that 'Member States shall ensure that official controls are carried out regularly, on a risk basis, and with appropriate frequency'. This means that risk managers should focus their resources on high risk entities: business operators, foodstuffs or particular hazards. During the risk-based planning of official controls, competent authorities have to take into account all of the objective evidence contributing to better decision making in the risk analysis framework. In this context, as a part of planning the most effective risk management options, different risk assessment and risk ranking methods are available, along with different planning techniques. Authorities have to choose whichever methods best fit their needs and resources. Continuous improvement and new methodologies are in the forefront of research.

Big Data and Network Science in the Field of Food Chain Safety

The need for handling, analysis and interpretation of large, interrelated datasets in various scientific fields, together with the rapid development of information-technology tools, have resulted in newly emerging data-related scientific fields. Their common characteristic is that with the use of computational science tools such rules or patterns could be identified which would otherwise be very hard or impossible using smaller datasets (Baranyi et al. 2013). Globalization, particularly its sociological and commercial aspects, started research of complex networks in the late 90's (Anderson and Marcouiller 2002). It quickly became evident that the structure and evolution of the networks showed many similarities regardless of what they represent (Baranyi et al. 2013). This phenomenon boosted research in different scientific fields which, after a short initial phase, network analysis methods found an application in many areas. It is used in sociology for the representation of the individuals and their relationships (Stanley and Katherine 1994; Salathé and Jones 2010), for mapping genes, proteins and their interactions with each other in molecular biology (Barabási and Albert 1999), and helps in the identification of business relationship of companies in different economical analysis (David and Douglas 1992).

As a definition (Börner et al. 2007) network science concerns itself with the study of different networks, be they social, biological, technological or scholarly networks. Its goal is to contrast, compare and integrate techniques and algorithms developed for a wide range of disciplines, primarily mathematics and statistics. Barabási (1999) compares the emergence of this science with sweeping developments in quantum mechanics in the 20th century. In his opinion, network sciences are building a theoretical and algorithmic framework which is energizing many research fields. "Born at the twilight of the twentieth century, network theory aims to understand the origins and characteristics of networks which hold together the components of various complex systems."

Data science, particularly network science, has an important role in food science enhancing security and safety of the food supply as well. Analysis tools based on network theory can be used in the risk-based control and monitoring systems of food business operators by analyzing their commercial relations with each other (Chmiel et al. 2007).

Borgatti et al. (2009) highlight that, over the past decades, network theory has supplied a valuable tool in explaining different social phenomena. In management science it has been widely used in supply chain management (Lazzarini et al. 2001), international trade analysis (Smith and White 1992), organizational development (Wasserman and Faust 1994) and policy analysis (Wagner and Leydesdorff 2005).

In the opinion of Fritz and Schiefer (2008), management challenges in closely co-operating enterprises, as well as the mutual dependence of all participants in the food chain, necessitate the application of network science in this area. The application of network science at the inter-firm level in agribusiness management is highlighted by Ng and Siebert (2009). In the most recent literature there are numerous examples of the successful application of network-theory approaches in the development of agribusiness systems, from the development of agricultural extension programs (Lehmann et al. 2012) to supply chain management (Farhat 2012).

Besides management science, network analysis is claimed to be an effective tool in food chain safety analysis as well. The first applications of network science in the field of food chain safety were aimed at mapping connections between countries or businesses. Petróczi et al. (2010, 2011) analyzed the notification data of the Rapid Alert System for Food and Feed (RASFF) of the European Commission. They identified European trade and notification patterns using network science methodology, using the model to forecast as well. Not only countries but different businesses were later analyzed for epidemiological purposes: Lentz et al. (2011) explored pig transport routes in Germany, showing hubs where cross infection was more likely. Ercsey-Ravasz et al. (2012) identified the most critical agri-food trade routes based on publicly available trade data. They drew attention to the fact that every second food batch produced is exported—and this proportion is increasing—providing proof of continuously growing international trade and an increasing need for the application of complex sciences.

The application of network theory for the analysis of animal migration has some decades-long tradition (Rommel et al. 1973; Harris 1979), but the conscious application of animal transportation data for the prevention of epidemiological problems is relatively new. This process is boosted by the rapid development of cattle identification systems. The comprehensive review of the New Zealand Ministry of Agriculture and Forestry offers a general overview on selected cattle identification and tracking systems worldwide (MAF 2009). The review proves, in which (1) all of the reviewed systems are implementing individual cattle identification requirements; (2) there is an increasing tendency to apply RFID technology; (3) most of them are mandatory; and (4) these systems are administered by governments or under industry-government partnerships. In the opinion of Schroeder and Tonsor (2012), cattle identification and traceability is becoming a necessary pre-condition for the international competitiveness of cattle and the cattle-product export market. In the last few years Dubé et al. (2008) has applied the network analysis approach to analyze and prevent foot and mouth disease. Martinez-López et al. (2009) have analyzed the trans-boundary flow of animals with the purpose of implementing disease prevention measures. Bajardi et al. (2011, 2012) mapped the Italian cattle trade network and made great progress in analyzing dynamic patterns, using network science tools to optimize cattle farm surveillance.

Motivation

When analyzing risk, food chain safety measures should be based on objective and scientifically based evidence. In most cases authorities are already using existing international risk assessments, risk ranking, risk-based priority setting tools, models, studies, and literature data. However, the data needed for substantiated risk assessment are in many cases not available. The lack of data or possible delays in providing updated records may hinder their use, especially for time-varying patterns (Valdano et al. 2015).

Our experience obtained in Hungarian and European food chain safety control planning systems show that, conventionally, risk is determined by the size of the herd as the frequency of official controls is determined by the number of animals present at any given time (taking into consideration other risk factors such as production type (dairy, meat, etc.) and the results of previous inspections). The calculations are based on the conventional risk approach as a product of severity of the hazard and the probability of occurrence. However, the effectiveness of this targeting mechanism can be questioned in many cases (Van Asselt et al. 2012).

One of the most important problems of this approach is that it doesn't take into account the network flow and the dynamics of the network; just the pure output or production data. The flow of animals denotes the animals transported from one node of the network to another during a given time period. A dynamic network is defined as a network where one or more of its relevant parameters (e.g. size of nodes, flow, etc.) changes as a function of time (Friesz et al. 1993).

The other drawback of the traditional risk based planning procedure is that it serves to set control priorities but is not suitable for epidemiological simulation exercises as the picture it captures is very static. Furthermore, the risk-based planning procedures of different member states are not cross-compatible, making international assessment very difficult or even impossible, and resulting in high coordination costs and significant delays when managing cross-border food chain incidents.

The cattle network consists of numerous closely cooperating holdings under the influence of natural (biological) and socio-economic factors, forming a network where the hubs of the network are the economic entities (e.g. farms, slaughterhouses, etc.) and the edges are the cattle-movements. The size of the hubs and edges can be considered as stochastic variables because the economic activities of the different entities show a considerable fluctuation. To minimize the risk of problems we have to understand the immanent structure of the network on the basis of network science. This will serve to fine-tune the strategy of decreasing risks of an epidemiological nature.

Network analysis is capable of capturing the time-dependent characteristics of the trade flow as well as selecting the highest risk nodes by their network characteristics. Furthermore, it is able to serve as a basis for epidemiological simulation exercises. Our motivation is to find a risk ranking tool which is able to capture those aspects of a functioning trade network.

Objectives

The aim of the present work is to develop a network-based assessment methodology, which is (1) suitable for the risk based planning of official controls (setting priorities based on network science) in order to place resources on the most critical elements of the cattle production network; (2) capable of simulating different epidemiological situations to increase preparedness for real epidemics using network-based spreading models. A majority of these models are based on system dynamics (Bagni et al. 2002) and in the last years agent-based simulation approach (Dion 2011), although there is a rapid development of Bayesian geo-statistical methods as well (Jewell et al. 2013; Ward et al. 2013). However, a critical point of all of these models is the quality of input data. Our results will shed light on how to prepare and interpret data for analysis. Finally, (3) to share the analysis methodology and algorithms with the network science and food chain safety community to enhance cross-compatibility of methods, making it possible to expand simulation exercises and risk-based planning processes across borders (since real world risks don't respect borders).

All of those objectives contribute to a methodology of letting decision-makers elaborate an optimized strategy for the inspection of cattle herds, control of cattle-traffic, and strategies for epidemiological crisis situations.

The hypotheses of the present work are (1) the Hungarian cattle network can be characterized as a scale free network; (2) the vulnerability of the network can be analyzed on the basis of the 'centrality' characteristics of different hubs of the network. The most vulnerable parts of the network are not necessarily the largest hubs, rather the ones which can be considered central parts of the network; (3) one centrality measure is not necessarily enough to characterize the centrality position of a given vertex, because the different centrality indicators reflect differently from every other aspect of the vertices (Friedkin 1991; Marsden 2002); (4) the Hungarian cattle network is a dynamic one. This means that the network size, the flow intensity of animals and other network properties, including centrality, can be characterized as considerably time-dependent.

We have applied wide ranging network analysis tools to determine the characteristic features of different nodes of the network and their time-variance. In this way we were able to characterize the Hungarian cattle trade network, improving the current control strategy and preparing for a possible crisis situation. This application of network science can be considered a relatively novel one as this paper shows a practical application of network theory by a food chain safety authority. This example of the application of a network science approach, based on big data analysis, can be considered a possible solution to a heretofore intractable big data-related problem in the food chain safety field.

Our aim was to shed light on the characteristic features of a given trade network (using the Hungarian cattle trade network as an example) for the purposes of increasing the effectiveness of food chain safety control and preparation for a possible outbreak. Lists of the most risky holdings obtained through the network analysis are used by risk managers while planning their annual control plans. Our model contributes to greater Hungarian food chain preparedness in a critical situation as it demonstrates a methodology which suitably determines the most critical parts of the network. It is not possible to offer a more concise and intelligible solution, because—as we will demonstrate – the actual features of the network are time-dependent variables.

Methods

Data Source

The cattle trade network is obtained using the database of the national cattle identification system (ENAR). This system is able to follow the animals along their whole life cycle from birth to slaughterhouse or from entering the territory of Hungary to their export. It has a legal background based on Regulation 1760/2000/EC on animal identification (European Parliament and the Council 2000), which makes the use of the system obligatory. In this way a continuous dataflow is generated, supplying more than 1000 lines of raw data each day. Each line represents an animal movement between two nodes. Each movement record reports the unique identifier of the animal, the codes of the holdings of origin and destination and the date of the movement.

To understand the structure and the characteristic features of this data flow an application of specific methods and approaches is needed. We have applied network analysis which has served as a tool for managing large amount of data.

Network analysis, as an interdisciplinary field of science, considers the relationship between organizations as a graph (Albert and Barabási 2002; Barabási 2002). The graph consists of a set of vertices and a set of edges (Tichy et al. 1979). In this case the vertices (or nodes) are (1) cattle-exporters to Hungary; (2) importers buying living cattle from Hungary and (3) various economic organizations—so called holdings—including farms, slaughterhouses, logistics/distribution centers, markets, artificial insemination stations, incinerators, fairs and animal health institutions. Movements are the transportation events of living cattle between different nodes. These are represented in our model as edges between nodes. These edges are called flows in graph theory when analyzing transportation processes (Wen and Arcak 2004).

The data inclusion criteria were: (1) time period between 01.01.2012.–31.12.2014.; (2) operating holdings with legal succession as well; (3) no limitation on age or birth of the animals (i.e. they didn't have to be born before 01.01.2012); (4) animals can die during the time period investigated; (5) all animals from the database.

The animal movements taken into consideration involved about 50,000 premises. The reason for the three-year time frame was to adequately characterize the Hungarian cattle-network for risk-based planning purposes yet not so much as to become outdated. The abovementioned trade routes represent approximately half a million movements a year. In the network, nodes may be active or inactive depending on whether farms sell or buy cattle in any given time frame.

The original raw data consisted of 4,667,479 lines, having 42,928,175 pieces of data altogether on animals and 713,482 on holdings. This static raw data was then cleaned and transformed through several steps into a static source-target matrix, containing data on 1,553,683 movements and 52,618 nodes. This static network was broken down into annual and monthly representations to analyze the behavior of the network over time. The basic network parameters were calculated each month resulting in a dynamic network containing 54,933,192 pieces of data attributed to nodes and 1,638,000 to edges.

This data-set can be described as a large-volume, complex, growing dataset concerning multiple, relatively autonomous parts. That's why it can be considered "big data" as defined by Wu et al. (2014), Power (2014), and Sonka (2014). The dataset satisfies the definition of the NIST group (2015) because it 'exceeds the capacity or capability of current or conventional methods or systems'. In the opinion of Ward and Barker, big data is not a set of data but 'a term, describing the storage and analysis of large and complex data steps using a series of techniques' (Ward and Barker 2013). The process of extracting insights from big data consists of five steps: (a) acquisition and recording, (b) extraction, cleaning and annotation, (c) integration, aggregation and representation, (d) modelling and analysis, (e) interpretation. Our current work contains all of these elements and lays down the basis for further modelling work.

Network Analysis Methods

The network of cattle holdings and movements were first analyzed to investigate the structure of the network and to calculate the main parameters. For each node the following measures were calculated:

The node ***degree*** is the number of relations (edges) of the nodes. However, in the case of the directed networks, the ***in-degree*** (incoming connections) and the ***out-degree*** (outgoing connections) values are important as well. Degree has generally been extended to the sum of weights when analyzing weighted networks and labelled node strength, so the ***weighted degree*** and the ***weighted in-*** and ***out-degree*** was calculated (Barrat et al. 2004; Newman 2001; Opsahl et al. 2010). These parameters offer an important piece of information on the intensity of relations between nodes and their environment. A high in-degree indicates that the node can be characterized as prominent; it receives animals from numerous farms. A high out-degree indicates that the node is influential because it has extensive connections with other farms. The same applies for weighted degrees but here the indicator shows not the number of connecting businesses, but the number of animals transported in and out.

In certain networks the nodes with the most important roles are the high degree nodes. However, this is a quite simplistic approach and, if the network has a strongly inhomogeneous structure (containing many clusters), it is certainly false. Low degree nodes connecting clusters in many cases play an important role in the network (Kleingberg 1999).

To understand the relative importance of different hubs in cattle flow besides the usual network metrics (e.g. in- and out-degrees, weighted degrees, etc.), we had to apply the centrality concepts of network analysis. Despite considerable research efforts invested into studying the centrality concept in network science, centrality is still an elusive concept which may be approximated from different perspectives where different centrality measures are available (Abbasi et al. 2012). We have analyzed betweenness centrality (Kim et al. 2012); closeness centrality (Freeman 1979); the two so-called prestige measures of centrality (Faust and Wasserman 1992): the hub centrality and the authority centrality (Kleinberg 1999, 2000), calculated using the HITS algorithm. There is a considerable difference between those centralities: In the case of the authority and hub centrality, a central node can be any node in the network, while in the case of betweenness and closeness centralities (as the names indicate) the central nodes cannot be the source-vertex or sink-vertex (Okoth and Wagner 2009). As defined by Newman (2005), a source vertex is a node with an in-degree zero while a sink vertex is a node without-degree zero.

The HITS algorithm was developed by Kleinberg (1999). This algorithm is a link analysis algorithm which helps in identifying the essential nodes in a graph. It consists of two scores, a ***hub score*** and an ***authority score***. The authority score of a node is a measure of the amount of valuable information that this node holds. The hub score of a node shows how many highly informative nodes or authoritative nodes this node points to. So a node with a high hub score shows that this node is pointing to many other authoritative nodes. On the other hand, a node with a high authoritative score shows that it is pointing to a large number of nodes, and as such, serves as a node of useful information in the network.

Betweenness centrality is an even more important statistical property of a network. This property is applied to a lot of real-world problems such as finding influential people in a social network, finding crucial hubs in a computer network, finding border crossing points which have the largest traffic or trade flow. The betweenness centrality of a node is an indicator of its centrality or importance in the network. It is described as the number of shortest paths from all the vertices to all the other vertices in the network that pass through the node in consideration (Brandes 2001).

Closeness centrality indicates how long it will take for information from a given node to reach other nodes in the network. The smaller the value, the more central role the node plays in the network.

We used Gephi open-source software for network visualization and analysis, making possible to use more than thirty algorithms and models. There are more than 100 plugins to the software, increasing the number of statistical tools (Devangana 2015). However, it was not necessary to apply these additional tools in our research. Further statistical analysis was made using Microsoft Excel software.

Results

Based upon network analysis it was possible to determine the most important (highest risk) flows in the system and construct different models for the cattle-network. On the basis of these models we have been able to determine the most important centers of the network which is extremely important because it is well-documented that the most vulnerable points of a network are not necessarily the largest hubs (Agarval et al. 2014; Wang et al. 2006; Wang et al. 2014).

The data-stream offered a possibility to determine the stability of different centralities of the system, as well as to analyze the stochastic relationships between these centrality indicators. As earlier stated, there is a considerable difference of importance between nodes according to their position in the network. The various statistical algorithms of the software tool provided a characterization of the cattle movement system, exploring both its structural and dynamical properties. There was an opportunity to compare these calculated values for each month and with this the central farms, logistics centers, slaughterhouses, and the peripheral holdings could be unveiled as well.

Network Structure

Mapping the Hungarian cattle holdings network, it was possible to calculate the basic metrics of the network. In network-related literature there is a wide range of indicators used to characterize a given network. Some of them aim to determine the position, sets and clusters of nodes and their connections. Another group of indicators describe the centrality of different nodes or offers information on network density. Other measures help to characterize the components, cores and cliques in the network. All of these pieces of information could furnish valuable insight into the network analyzed but we had to limit ourselves to simple characteristic features of the network.

A key property of each node (in this case, holdings) is its *degree*, representing the number of links it has to other nodes. In the cattle network it means the number of business partners. The *degree distribution*, p_k, provides the probability that a randomly selected node in the network has degree k. For a network with N nodes the degree distribution is given by the equation:

$$1)\qquad p_k = \frac{N_k}{N}$$

where N_k is the number of degree-k nodes. The degree distribution has a very important role in network theory following the discovery of scale-free networks (Barabási and Albert 1999). The degree distribution of the Hungarian cattle holdings network (Figure 1**)** shows a very characteristic heavy-tailed distribution specific to *scale-free* networks.

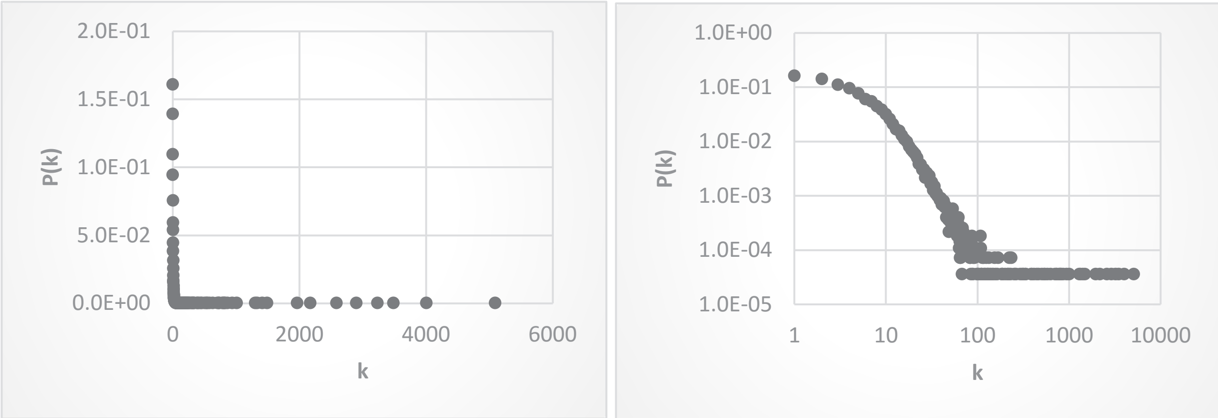


Figure 1. Different plots of the degree distribution of the Hungarian cattle holdings network (timeframe: 2012–2014); linear plot (left), log-log plot (right).

This heavy-tailed distribution shows that there are many small nodes (with few connections), and there are few very large nodes (with a lot of connections). The scale-free networks are networks whose degree distribution follows a power law. To prove the power law distribution and to obtain the degree exponent (γ), which is important for further analysis, a cumulative distribution was plotted and then a power law curve was fitted (Figure 2).

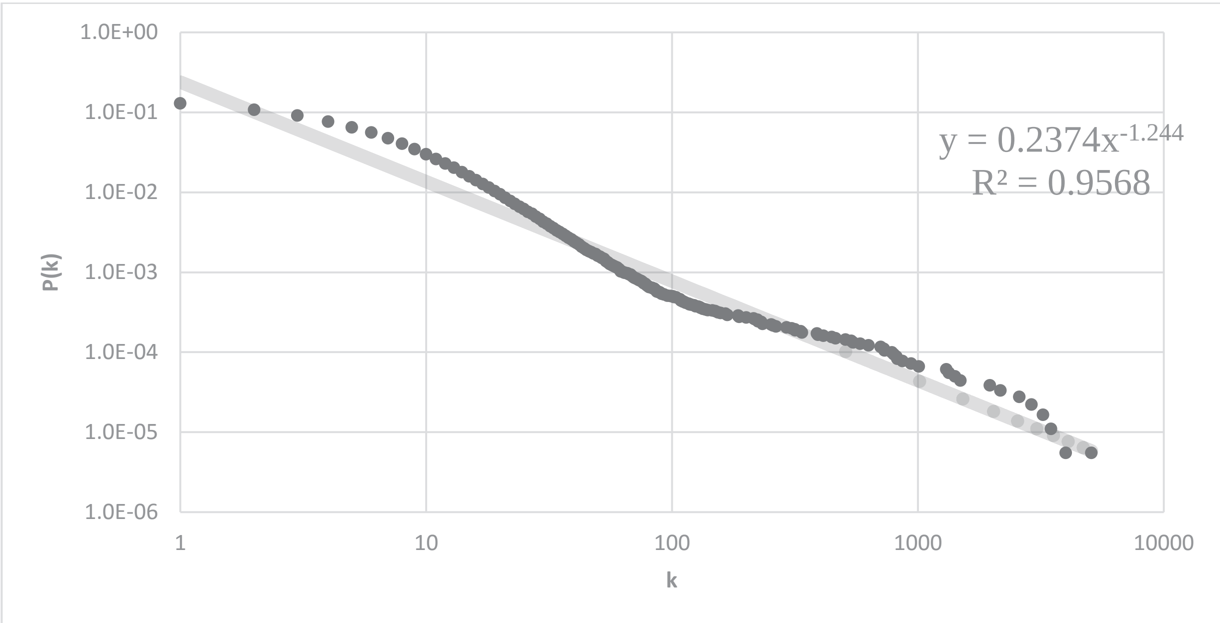


Figure 2. The degree distribution of the Hungarian cattle holdings network presented as a cumulative log-log plot from 2012–2014.

2) $$P_k = \sum_{q=k+1}^{\infty} p_q$$

In case of power law the cumulative distribution scales as

3) $$P_k \sim k^{-\gamma+1}$$

The degree exponent for the Hungarian cattle holdings network is **2.24**. As this network is a directed network, the scale-free property applies separately to the in- and the out-degrees (Figure 3).

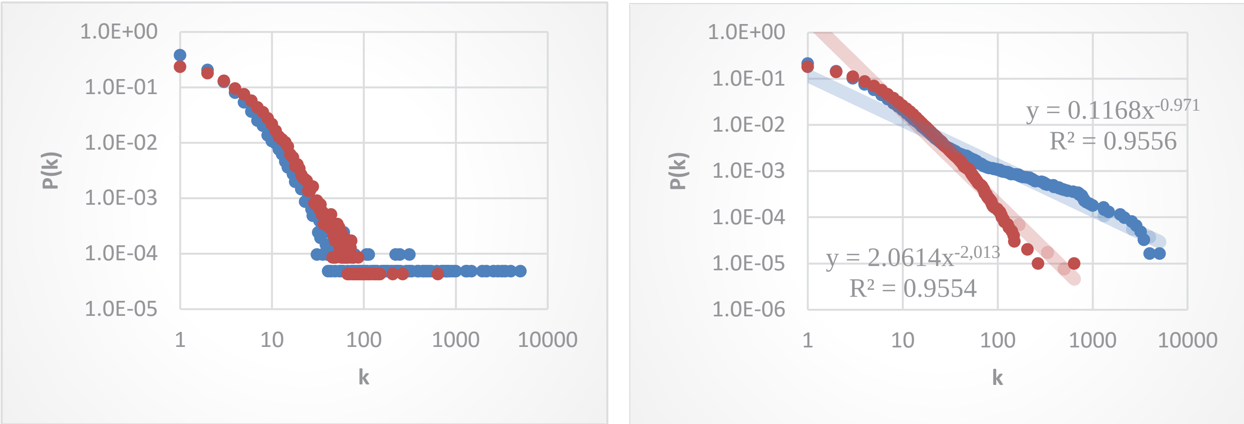


Figure 3. The in-degree and out-degree distributions of the Hungarian cattle holdings network from 2012–2014.

Note. Represented as a log-log plot (left) and a cumulative log-log plot (right). The in-degree is marked with blue and the out-degree is marked with red dots.

As it can be observed, the degree exponents are different for in-degree ($\gamma = 1.97$) and out-degree ($\gamma = 3.01$), showing a substantial difference between the two. This is attributable to the specific nodes with a very high in-degree e.g. slaughterhouses.

Similarly, the degree distribution can be calculated for the weighted degrees as well. Weighted degrees represent the size of the traffic going through a node; in this case, number of animals transported to and from the holdings (Figure 4.). The traffic size is important for food chain safety reasons because any epidemiological problem in a herd with intense traffic can be proliferated in the network extremely rapidly. That's why this piece of information helps risk management. The degree exponents can be calculated from the cumulative log-log plots of the weighted degree distribution (Figure 5).

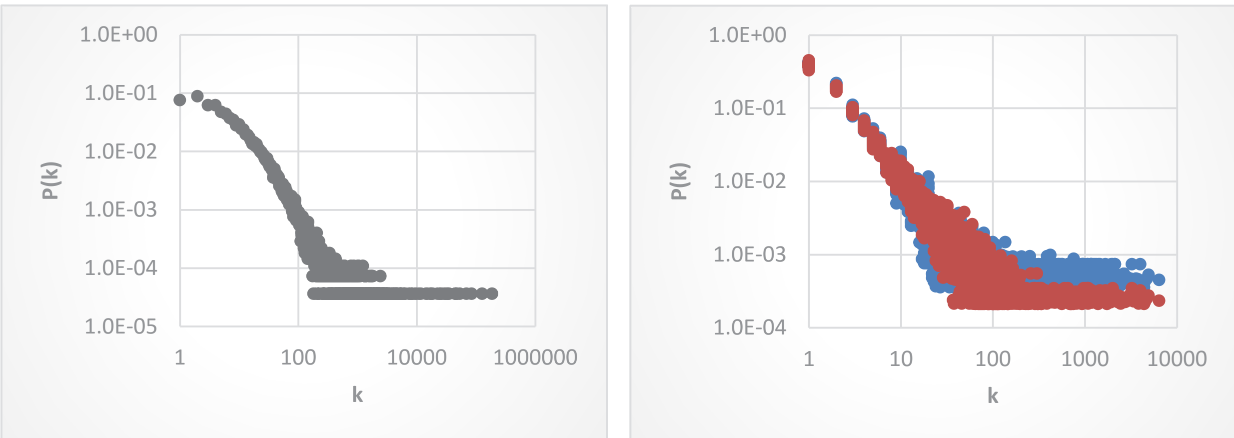


Figure 4. The weighted degree distributions of the Hungarian cattle holdings network represented as log-log plots from 2012–2014.

Note. The graphs show weighted degree (left) and a weighted in- and out-degree (right) distributions. The weighted in-degree is marked with blue and the out-degree is marked with red dots.

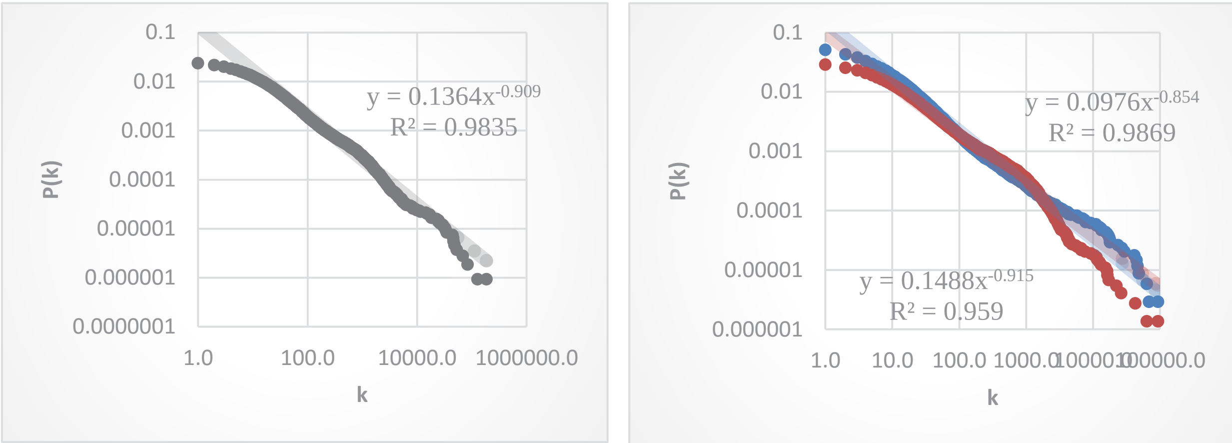


Figure 5. Weighted degree distributions of the Hungarian cattle holdings network represented as cumulative log-log plots from 2012–2014.

Note. The graphs show a weighted degree (left) and a weighted in- and out-degree (right) distributions. The weighted in-degree is marked with blue and the out-degree is marked with red dots.

Interestingly, degree exponents of the weighted degree, weighted in- and out-degree distributions are very similar to each other and fall in the range of $\gamma \sim 1.9$. The reason behind it is that the large weighted degree nodes are typically logistic centers where the incoming and outgoing flows are identical.

The results above show that this network has all the intrinsic properties of other scale-free networks, highlighting the fact that some holdings have a critical role in the network. Identifying those, we make a step towards controlling them. The scale-free name captures the lack of an internal scale, a consequence of the fact that nodes with widely different degrees co-exist in the same network. This feature distinguishes scale-free networks from lattices, in which all nodes have exactly the same degree ($\sigma = 0$), or from random networks whose degrees vary in a narrow range ($\sigma = \langle k \rangle^{1/2}$). This divergence is the origin of some of the most intriguing properties of scale-free networks, from their robustness to random failures to the anomalous spread of viruses (Barabási 2015).

This means that this network is quite robust against random failures but vulnerable in case of targeted attacks.

Having a degree exponent between two and three means this network also shows small world properties—meaning that only a few steps are needed to get from a random point to another random point—having an important implication in the case of spreading diseases. The average path length (steps needed to reach any random node from any other random node) for the Hungarian cattle holdings network is 6.92 for the three-year period.

Basic Network Properties

Devising the network's basic structural properties, other valuable information could be extracted from other network measures or indicators. As specified earlier, the degree, in-degree, out-degree, weighted degree, weighted in-degree and weighted out-degree of the nodes was calculated for different holdings, having an objective to set priority lists for different control purposes. The top five nodes (highest risk nodes) listed according to various properties are presented in Table 1.

Table 1. Top five nodes based on degree, in-degree, out-degree, weighted degree, weighted in-degree and weighted out-degree (timeframe: 2012–2014).

	ID	Type	Degree		In-Degree		Out-Degree		Weighted Degree		Weighted In-Degree		Weighted Out-Degree	
			Rank	Value	Rank	Value	Rank	Value	Rank	Value	Rank	Value	Rank	Value
Top 5 Degree	528781	L	1	5093	1	5091	16781	2	2	127196	3	63601	2	63595
	720342	S	2	4007	2	4007	49104	0	6	48419	4	48419	49104	0
	919273	L	3	3489	3	3483	6779	6	1	188300	1	94153	1	94147
	242344	S	4	3235	4	3233	16911	2	9	44956	6	44951	14995	5
	257710	S	5	2902	5	2901	20045	1	14	30050	8	30046	16035	4
Top 5 In-Degree	528781	L	1	5093	1	5091	16781	2	2	127196	3	63601	2	63595
	720342	S	2	4007	2	4007	49104	0	6	48419	4	48419	49104	0
	919273	L	3	3489	3	3483	6779	6	1	188300	1	94153	1	94147
	242344	S	4	3235	4	3233	16911	2	9	44956	6	44951	14995	5
	257710	S	5	2902	5	2901	20045	1	14	30050	8	30046	16035	4
Top 5 Out-Degree	230129	M	10	1417	19	774	1	643	55	4301	46	2503	112	1798
	456485	M	24	583	34	316	2	267	241	1362	114	764	392	598
	357532	F	48	217	1342	10	3	207	78	3099	1945	38	43	3061
	355807	M	35	318	52	167	4	151	370	890	149	521	556	369
	145593	F	60	150	9323	2	5	148	215	1472	2575	29	162	1443
Top 5 Weighted Degree	919273	L	3	3489	3	3483	6779	6	1	188300	1	94153	1	94147
	528781	L	1	5093	1	5091	16781	2	2	127196	3	63601	2	63595
	860928	F	20	734	21	725	3444	9	3	83254	7	39662	3	43592
	490540	I	7	2172	7	2168	10228	4	4	68908	2	68904	16316	4
	217330	F	39	265	39	247	930	18	5	54681	9	27528	4	27153
Top 5 Weighted In-Degree	919273	L	3	3489	3	3483	6779	6	1	188300	1	94153	1	94147
	490540	I	7	2172	7	2168	10228	4	4	68908	2	68904	16316	4
	528781	L	1	5093	1	5091	16781	2	2	127196	3	63601	2	63595
	720342	S	2	4007	2	4007	49104	0	6	48419	4	48419	49104	0
	582987	S	6	2590	6	2588	16542	2	8	45170	5	45153	6869	17
Top 5 Weighted Out-Degree	919273	L	3	3489	3	3483	6779	6	1	188300	1	94153	1	94147
	528781	L	1	5093	1	5091	16781	2	2	127196	3	63601	2	63595
	860928	F	20	734	21	725	3444	9	3	83254	7	39662	3	43592
	217330	F	39	265	39	247	930	18	5	54681	9	27528	4	27153
	941088	F	52	187	50	170	1077	17	7	47053	10	23801	5	23252

Note. The holding IDs are anonymized. L = logistics/distribution center; S = slaughterhouse, M = animal market; F = farm; I = incinerator.

As can be derived from the results presented in the table, the different network properties have different meaning from a real-life control perspective. Degree shows the connections between different holdings. The stability of these connections have implications on risk, in line with human epidemiology (e.g. in case of sexually transmitted diseases spread). The holdings tending to be more loyal to their business partners have lower risks compared to those switching their partners over time (Valdano et al. 2015).

The considerable (one order of magnitude) differences between the in- and out-degree values and distributions hold important information as well: nodes with the highest in-degrees (and weighted in-degrees) are slaughterhouses. They don't release animals so their out-degree is (virtually) 0; they are so-called sink vertices. In the case of hygiene, documentation, or traceability controls they are very important hubs to control and, in the case of some diseases, large slaughterhouses may be places where cross-contamination or cross-infection occurs. However, in case of other diseases they don't represent a real high risk vertex and they should be excluded from analyses since this is an end-point to animals. This implies that, depending on the actual control objective and the characteristics of the causative agent, those nodes should be included or excluded from analysis on a case-by-case basis.

Similarly, from a control perspective, it is important to observe nodes with low degree - high weighted degree (high trade activity with limited number of business partners), or high-degree, relatively low weighted degrees (high number of business partners, but limited trade with each of them), usually being markets or trans-loading stations.

As it is shown in the table, the maximum in-degree value of the network is 5091 and it belongs to a slaughterhouse; while the maximum out-degree value is only 643 (a market), confirming the phenomenon observed during degree distribution analysis, resulting in different degree exponents for in- and out-degrees. In contrary, the maximum values of the weighted in- and out-degrees are similar: 93,663 and 93,657, respectively (logistic center).

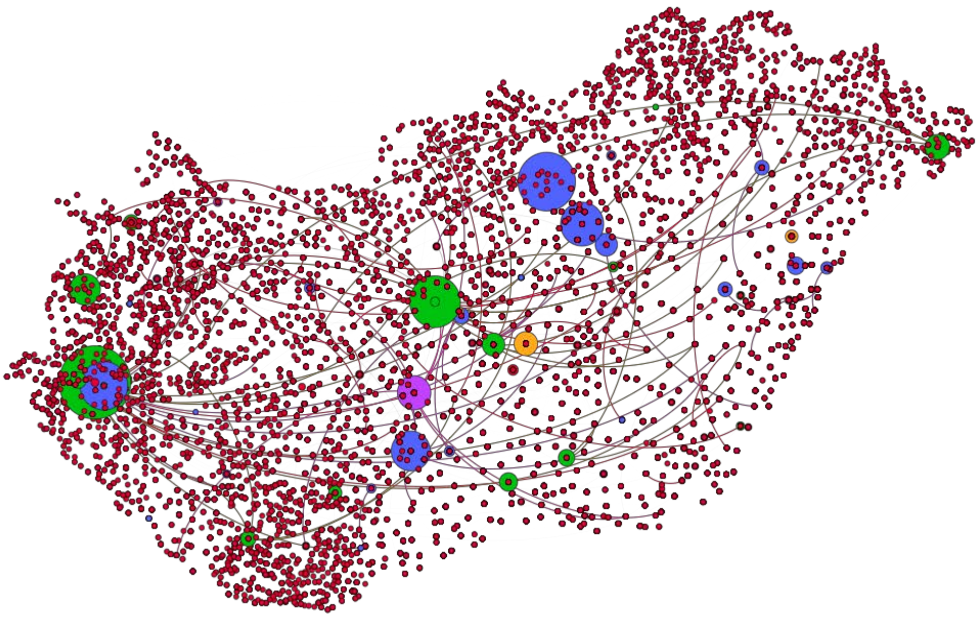


Figure 6. Geographical representation of the Hungarian cattle holdings network.

Note. The size of the nodes is influenced by degree, the color depends on the production type (red = livestock farm; blue = slaughterhouse; green = logistics/distribution center; orange = market; purple = incinerator). The color of the edges is influenced by source node, and the weight is limited to minimum 36 (at least one movement per month on average).

Pay attention to the fact, that 'export' was part of the dataset as one single node (since there is no information about the exact recipient holding), and the connections contributed to the degree values of the nodes but it was excluded from the ranking exercise. If we knew about the actual destination of the exported cattle, the out-degree would increase (with unchanged weighted out-degree). Similarly, the source of imported animals was present in the analysis as single nodes for each exporting country. Having information about the exact source holdings, the in-degree values would increase (with unchanged weighted in-degree). The export activity is far larger than import (207,094 and 55,240 movements in the three-year period, respectively), meaning if more precise data on holdings outside Hungary were present the difference between in- and out-degree distribution would decrease. The geo-layout of the Hungarian cattle holdings network is presented on Figure 6.

Centrality Measures

The most vulnerable points of a network are not necessarily their largest hubs, as discussed previously. To extract information on the nodes playing a central role in the network, different centrality measures were calculated: betweenness, closeness, authority and hub centralities were determined. The vertices of high betweenness centrality value are usually logistic centers, transloading places or major livestock farms. These nodes have an important role in epidemiological investigations because of the high risk of cross-infections. The top five nodes sorted according to different centrality values are presented in Table 2.

Table 2. Top five nodes based on betweenness centrality, closeness centrality, authority and hub centrality (timeframe: 2012-2014).

	ID	Type	Betweenness centrality		Closeness centrality		Authority		Hub centrality	
			Rank	Value	Rank	Value	Rank	Value	Rank	Value
Top 5 Betweenness centrality	230129	M	1	124525786	1	4,353	3	0,00529509	3	0,00658296
	769670	F	2	87082032	21	4,971	8	0,00267829	8	0,00293686
	456485	M	3	74391523	2	4,481	9	0,00216586	9	0,00255726
	919273	L	4	44847391	1893	6,118	1	0,02380399	1	0,02744064
	447999	F	5	34191545	252	5,285	14	0,00121616	12	0,00143846
Top 5 Closeness centrality	230129	M	1	124525786	1	4,353	3	0,00529509	3	0,00658296
	456485	M	3	74391523	2	4,481	9	0,00216586	9	0,00255726
	448931	F	40	3803558	3	4,640	957	0,00008199	638	0,00011987
	583058	F	39	3893053	4	4,761	3496	0,00003416	2908	0,00004995
	806192	F	87	2191809	5	4,817	3499	0,00003416	2911	0,00004995
Top 5 Authority	919273	L	4	44847391	1893	6,118	1	0,02380399	1	0,02744064
	490540	I	6	29500637	3143	6,530	2	0,01481942	2	0,01512382
	230129	M	1	124525786	1	4,353	3	0,00529509	3	0,00658296
	860928	F	44	3660243	2597	6,362	4	0,00494664	4	0,00569391
	431898	F	9	18940054	249	5,277	5	0,00335470	5	0,00379594
Top 5 Hub centrality	919273	L	4	44847391	1893	6,118	1	0,02380399	1	0,02744064
	490540	I	6	29500637	3143	6,530	2	0,01481942	2	0,01512382
	230129	M	1	124525786	1	4,353	3	0,00529509	3	0,00658296
	860928	F	44	3660243	2597	6,362	4	0,00494664	4	0,00569391
	431898	F	9	18940054	249	5,277	5	0,00335470	5	0,00379594

Note. The holding IDs are anonymized. L = logistics/distribution center; S = slaughterhouse, M = animal market; F = farm; I = incinerator.

On the basis of the centrality measure other extremities–the "peripheral holdings" could be defined. These entities are not regular participants of the global cattle network. Their role is marginal in the network as a whole, but—taking into consideration their often low technological level—it is important to include their activity because they can be sources of epidemiological problems.

As described earlier, different centrality concepts capture different aspects of a central node. It is an important question from a food chain safety perspective which concept is the most useful from a risk analysis point of view. It is out of the scope of this paper to answer this question. However, with the help of the large amount of data it was possible to analyze the stochastic relationships between the centrality indicators. We decided to filter our analysis since, given all the nodes from the network, correlation figures are largely biased due to sink vertices, export (represented as one single node) and import (source countries as nodes) data and the holdings characterized by small throughput. Therefore, the nodes (and the corresponding edges) outside Hungary were excluded from the calculation of centrality values, then the nodes with <3 in-degree and out-degree (at least one in and out connection a year) were excluded from the correlation analysis as well as nodes with betweenness centrality value of 0 and closeness centrality value of 1 (nodes with a small number of connections, not being part of the giant component of the network). Then the correlation between the different centrality results was calculated (Table 3)**.** To understand the relationship between different centrality indicators, we have applied regression analysis. This is an extremely important step because on this basis we will be able to understand whether there is a possibility to decrease the number of centrality indicators to judge the position of a given vertex of not.

As seen from the results, there is a weak negative correlation between closeness and betweenness centralities, and a stronger correlation between betweenness centrality and hub centrality and authority. The strong relationship between hub and authority centrality can be explained by their similar role: in the opinion of Kleinberg (1999) hubs and authorities stand in a mutually reinforcing relationship. Valente et al. (2008) showed in their work a slight correlation between betweenness and closeness centrality, indicating that these measures are distinct, yet conceptually related.

Table 3. Stochastic relationship between betweenness centrality, closeness centrality, authority and hub centrality in case of the Hungarian cattle holdings network.

	Betweenness centrality	**Closeness centrality**	**Authority**
Betweenness centrality	1		
Closeness centrality	-0.1263	1	
Authority	0.4932	-0.0619	1
Hub centrality	0.5054	-0.0634	0.9975

Note. Time period: 2012–2014; directed network

Dynamic Patterns

The dataset offered a possibility to analyze the dynamic patterns of the network, to observe and draw conclusions on the time-dependent features which may have an influence on the planning of control activities. To that aim, monthly, annual and the whole dataset for three years were compared in this section.

The simplest approach is to observe the number of animals moving per month. The results (Figure 7) indicated that the trade becomes very active in June–July with a peak of activity at the end of the year. The trend of increasing activity in the second half of the year seems to be stable. This should have an impact on the control time schedules, assigning increased control frequencies to those periods.

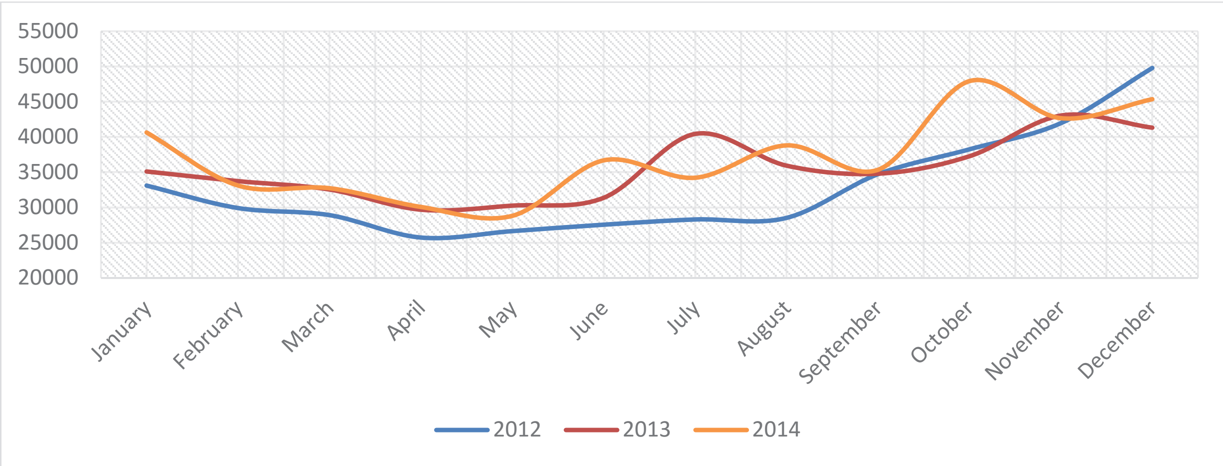


Figure 7. Monthly movement of the animals in the Hungarian cattle holdings network.

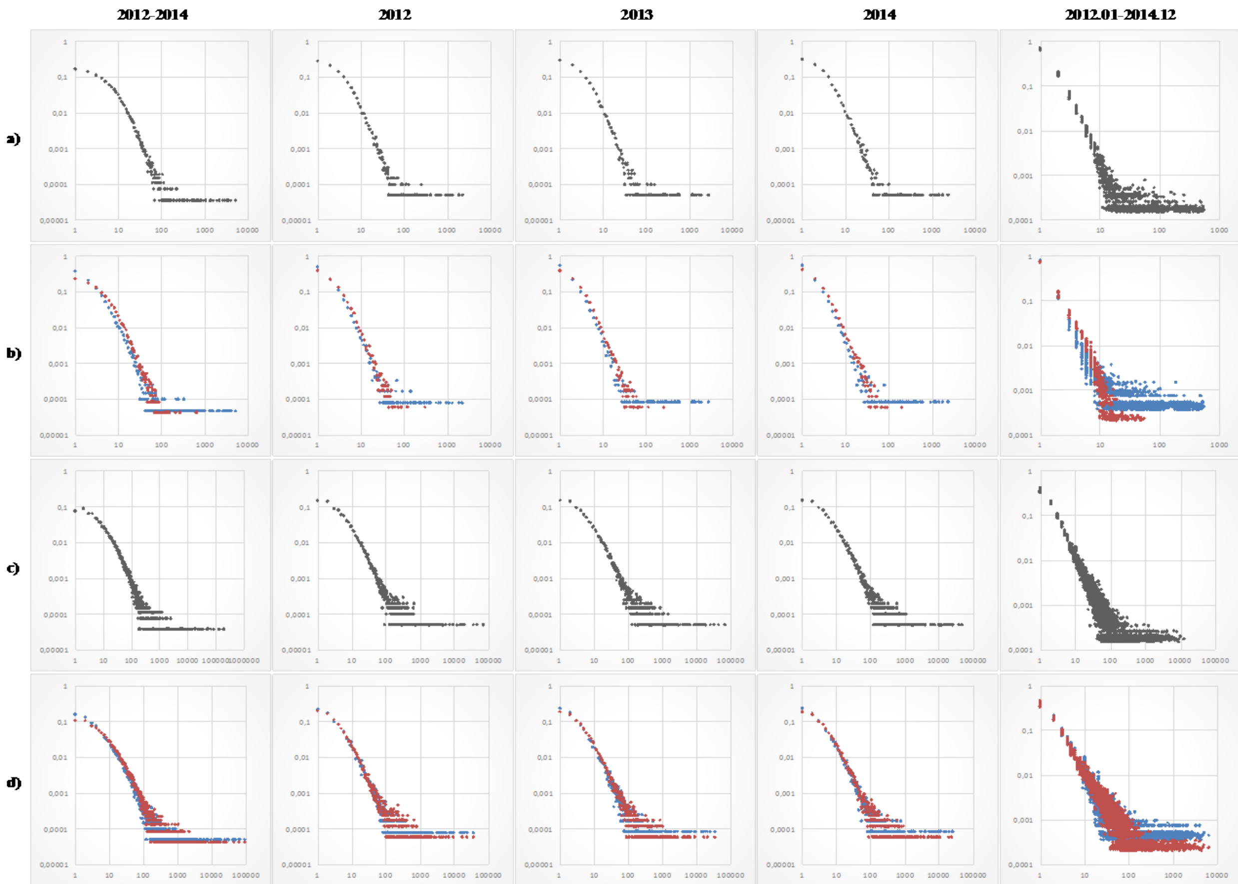


Figure 8. Dynamic patterns of the Hungarian cattle holdings network.

Note. a) Degree; b) In-Degree (blue) and Out-Degree (red); c) Weighted Degree; d) Weighted In-Degree (blue) and Weighted Out-Degree (red) distributions presented in 3-year, annual and monthly breakdowns.

During analysis, the changes in the activity of the holdings were recorded along with the analysis of the dynamic patterns of the entire network properties (Figure 8).

It can be derived from the results that, apart from small differences, network characteristics are quite stable over time, allowing for predictions at the overall network level. We selected the top five nodes on the betweenness centrality rank list and plotted the monthly betweenness centrality values to see when the nodes played a central role in the network over the three year period. The results (Figure 9) show a very volatile nature (differences of many orders of magnitude between months) of the holdings in relation to betweenness centrality values. It also shows that the analysis of the dynamic patterns is valuable especially in case of single holding analysis: performing time-dependent assessments, the results could be used for effective targeting of control, or prediction purposes as well.

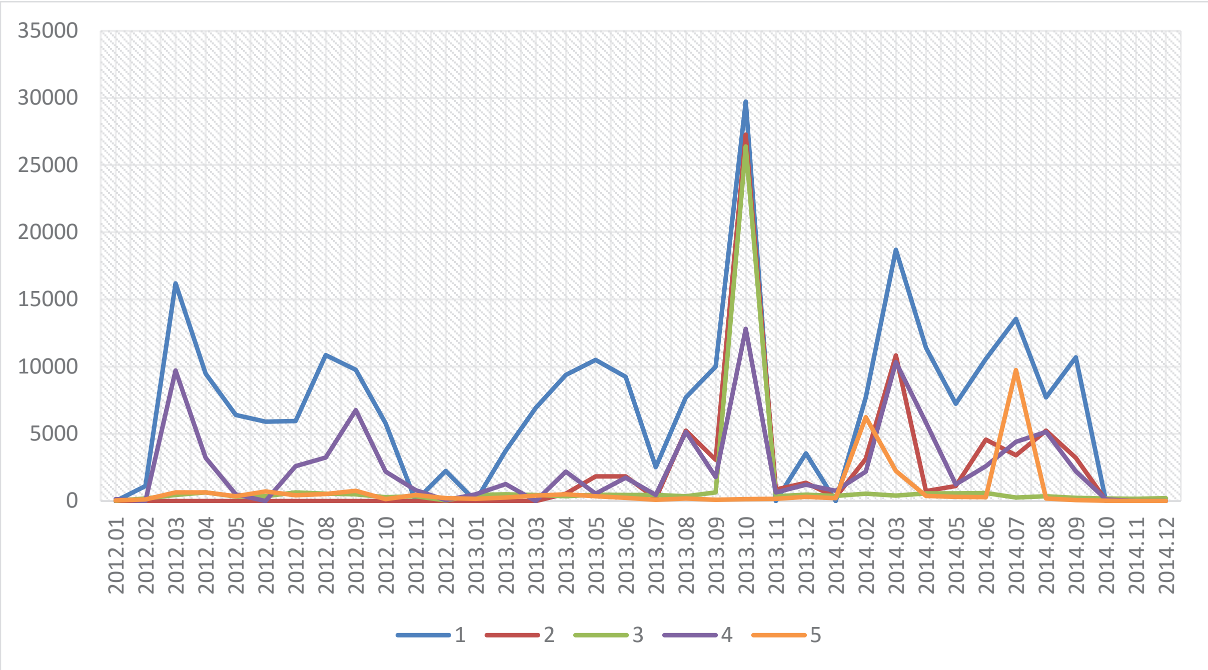


Figure 9. Changes in betweenness centrality values of the top five holdings of the Hungarian cattle holdings network during a thirty-six- month period.

Application of the Results in Practice

The results presented above all contribute to network analysis based, risk based control plans. The outputs of the analysis served as valuable input information in the planning process of official control plans. As an output, 100 highest risk cattle holdings were selected for food chain safety control based on the basic network properties, as well as on centrality measures. Those holdings are controlled for biosecurity measures, hygiene, animal welfare rules, safety assurance systems, documentation, etc. Furthermore, 100 highest risk holdings were selected for animal identification control. Those holdings are controlled for the identification and traceability rules.

The analyses performed provide information on the source and routes of possible infections so that preventive and control measures can be applied, increasing preparedness of the food chain stakeholders. In case of an outbreak, the mapped network makes a rapid traceability and epidemic spreading prediction possible, allowing for effective risk management.

Implications

Globalization, the data explosion, fast changing trade routes and food technologies are the important drivers which inspire us to develop new analysis and assessment methods in the field of food chain safety. There is a strong need for an interdisciplinary approach to monitor, understand, and control the trade-flow in the food chain.

The data on movements of cattle are increasingly becoming available thanks to identification and tracing systems put in place in the European Union. By using the approaches and techniques of network science it is possible to analyze the dynamic system of cattle movements, going beyond static and simple approximations (Bajardi et al. 2011; Natale et al. 2009).

During the analysis of the Hungarian cattle holdings network the basic structure of the network was revealed, showing scale-free properties, thus having serious implications from a food chain safety control perspective. This network has small world properties meaning that – because of the hubs and high centrality nodes—there is a small distance between any random holdings, potentially resulting in a rapid spread of epidemic as the spread of a pathogen on a scale-free network is instantaneous (Barabási 2015). This should be taken into account during the preventative measures (including e.g. vaccination strategies) at a business level as well as at an official control planning level. Furthermore, this phenomenon has very important implications for the forecasting and risk or crisis management in case of an actual outbreak, showing the possible advantages of using network spreading models in conjunction with traditional epidemiological modelling (Pastor-Satorras and Vespignani 2001).

The other consequence of having scale-free properties is the vulnerability to intentional attacks against the hubs or central nodes, showing the growing need for the control and preparedness of that critical infrastructure. It is important to emphasize that intentional attacks on the network behave differently and need a slightly different analytical and risk management approach compared to unintentional events. Epidemics follow the rules set by the characteristics of the infectious agents, while in case of intentional attacks, different spreading models should be used, based on socio-psychological and economic analysis.

On this basis, suggestions have been formulated for the food chain safety authority determining which farms should be the focus of their control activity. The list of the highest risk holdings obtained by network analysis is used directly by risk managers when outlining their annual control plans. Should any epidemiological problem occur, the easily updatable database on network characteristics offers essential input for further optimization of the control strategy. The tool used is suitable for a rapid assessment of a huge and complex system within minutes after data cleaning. It is possible to give a very informative graphical representation of the cattle holding network, making possible to easily choose control or audit targets. During the analysis of the dynamic properties of the network we revealed further possibilities to explore, hence making network based epidemiological simulations the next item on our research agenda.

As it could be seen from the difference of the 'real life' meaning of various network properties, and considering the implications of those, critical thinking during the application of the results is essential. Substantial knowledge of the food chain safety science is needed for the correct interpretation of the network analysis results and advanced skills in computational science are important in extracting valuable information from the underlying network data. This inter- and multidisciplinary field of science calls for such experts and the need for capacity building.

In addition, the software tools available are not by themselves suitable alone network-based food chain safety analyses. The data cleaning, transformation and enrichment steps needed to obtain a dynamic dataset suitable for network-analysis (as demanded by the software used) require many steps and, after the usual network analysis, many calculations are done over using other analysis tools. This calls for dedicated software development in the future, to decrease the time needed between receiving raw data and delivering pertinent information to decision makers. In epidemic situations, time is of utmost importance.

An important aim of this ongoing research is to share the methodology and algorithms with the network science and food chain safety community, in order to enhance the capacity building process and to improve the cross-compatibility of the methods. This makes it possible to expand simulation exercises and risk based planning processes across borders, as real world situations don't respect borders either. For that reason, the anonymized raw data, the data cleaning process, the analysis algorithms and the Gephi software settings used are published on the website of Hungarian National Food Chain Safety Authority (NÉBIH).[1] Furthermore, for the sake of better illustration, particularly for educational purposes, the key issues of the article (graph-dynamics) are illustrated in a Prezi, based on a series of Gephi files on the same site.

This study opens the road to future work in several directions. This work contributes to 1) determining the most vulnerable parts of a cattle holding network; 2) increasing the effectiveness of the control of the cattle-flow; 3) revealing the interdependencies; 4) helping to work out an optimized strategy for the inspection of herds; 5) increasing the preparedness against outbreaks and intentional attacks; 6) enhancing epidemiological modelling simulations; 7) providing information on the source of possible infections so that preventive and control measures can be applied; and finally 8) serving the food chain safety and network science community with analyzable data and helpful descriptions of the methodology to enhance cross-border co-operation.

References

Abbasi, A., L. Hossain, and L. Leydesdorff. 2012. Betweenness centrality as a driver of preferential attachment in the evolution of research collaboration networks. *Journal of Informetrics* 6 (3): 403–412.

Agarwal, P. K., S. Har-Peled, H. Kaplan, and M. Sharir. 2014. Union of random Minkowski sums and network vulnerability analysis. *Discrete & Computational Geometry* 52 (3): 551–582.

Albert, R. and A.-L. Barabási. 2002. Statistical mechanics of complex networks. *Reviews of Modern Physics* 74 (1): 47–91.

Anderson, J. E. and D. Marcouiller. 2002. Insecurity and the Pattern of Trade: An Empirical Investigation. *Review of Economics and Statistics* 84 (2): 342–352.

Bagni, R., R. Berchi, P. Cariello, 2002. A comparison of simulation models applied to epidemics. *Journal of Artificial Societies and Social Simulation* 5 (3).

[1] http://portal.nebih.gov.hu/web/guest/-/network-science-based-decision-support-in-food-chain-safety-systems.

Bajardi, P., A. Barrat, F. Natale, L. Savini, and V. Colizza. 2011. Dynamical Patterns of Cattle Trade Movements. *PLoS ONE* 6 (5): e19869.

Bajardi, P., A. Barrat, L. Savini and V. Colizza. 2012. Optimizing Surveillance for Livestock Disease Spreading Through Animal Movements. *Journal of the Royal Society Interface* 9 (76): 2814–2825.

Barabási, A.-L. 2002. *Linked: The New Science of Networks*. Perseus Books Group, USA. ISBN 0-452-28439-2.

Barabási, A.-L. 2015. Network Science. http://barabasi.com/networksciencebook/ [accessed October 20, 2015]

Barabási, A.-L. and R. Albert. 1999. Emergence of Scaling in Random Networks. *Science* 286 (5439): 509–512.

Baranyi, J., A. Jóźwiak, L.Varga, M. Mézes, J. Beczner, and J. Farkas. 2013. [Application Potentials of Network Science, Bioinformatics and Systems Biology to Food Science]. *Hungarian Science* 174 (9): 1094–1102.

Barrat, A., M. Barthélémy, R. Pastor-Satorras, and A.Vespignani, 2004. The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences* 101 (11): 3747–3752.

Bojnec, S. and I. Fertő. 2009. Agro-food trade competitiveness of Central European and Balkan countries. *Food Policy* 34 (5): 417–425.

Bökönyi, S. 1971. The Development and History of Domestic Animals in Hungary: The Neolithic Through the Middle Ages. *American Anthropologist* 73 (3): 640–674.

Borgatti, S. P., A. Mehra, D. J. Brass, G. Labianca. 2009. Network Analysis in the Social Sciences. *Science* 323 (5916): 892-895.

Börner, K., S. Sanyal, A. Vespignani. 2007. Network science. *Annual Review of Information Science and Technology* 41 (1): 537-607.

Brandes, U. 2001. A Faster Algorithm for Betweenness Centrality. *Journal of Mathematical Sociology* 25 (2): 163–177.

Chmiel, A. M., J. Sienkiewicz, K. Suchecki and J. A. Hołyst. 2007. Networks of companies and branches in Poland. *Physica A: Statistical Mechanics and its Applications* 383 (1): 134–138.

Coulombier, D. and J. Takkinen. 2013. From national to international challenges in cross-border multi-country, multi-vehicle foodborne outbreak investigations. *Euro surveillance: bulletin Européen sur les maladies transmissibles = European communicable disease bulletin* 18 (11): 20423.

Csáki, C. 1990. Agricultural Changes in Eastern Europe at the Beginning of the 1990s. *American Journal of Agricultural Economics* 72 (5): 1233–1242.

Csizmadia, E. 1974. New features in Hungarian agriculture in the 1970s. *Acta Oeconomica* 12 (3-4): 349–365.

David, A. S. and Douglas, R. W. 1992. Structure and Dynamics of the Global Economy: Network Analysis of International Trade 1965–1980. *Social Forces* 70 (4): 857–893.

Devangana, K. 2015. *Gephi Cookbook*. Packt Publishing. Birmingham, UK. ISBN: 978-1-78398-741-2.

Dion, E., L. VanSchalkwyk, E.F. Lambin, 2011. The landscape epidemiology of foot-and-mouth disease in South Africa: A spatially explicit multi-agent simulation. *Ecological Modelling* 222 (13): 2059–2072.

Dubé, C., C. Ribble, D. Kelton, and B. McNab. 2008. Comparing network analysis measures to determine potential epidemic size of highly contagious exotic diseases in fragmented monthly networks of dairy cattle movements in Ontario, Canada. *Transboundary and Emerging Diseases* 55 (9-10): 382–392.

Dubé, C., C. Ribble, D. Kelton, and B. McNab. 2009. A Review of Network Analysis Terminology and its Application to Foot-and-Mouth Disease Modelling and Policy Development. *Transboundary and Emerging Diseases* 56 (3): 73–85.

Ercsey-Ravasz, M., Z. Toroczkai, Z. Lakner, and J. Baranyi. 2012. Complexity of the International Agro-Food Trade Network and Its Impact on Food Safety. *PLoS ONE* 7 (5): e37810.

FAO/WHO. 2007. *Food safety risk analysis. A guide for national food safety authorities.* Food and Agriculture Organization of the United Nations, Rome, Italy. http://www.fao.org/docrep/012/a0822e/a0822e00.htm [accessed October 5, 2015]

Farhat, R., 2012. Management of logistics and ICT in food supply chain network: a conceptual framework. *International Journal of Logistics Economics and Globalisation* 4 (3): 162-178.

Faust, K. and S. Wasserman. 1992. Centrality and prestige: A review and synthesis. *Journal of Quantitative Anthropology* 4 (1): 23–78.

Freeman, L. C. 1979. Centrality in social networks conceptual clarification. *Social Networks* 1 (3): 215–239.

Friedkin, N. E. 1991. Theoretical foundations for centrality measures. *American Journal of Sociology* 96 (6): 1478–1504.

Friesz, T. L., D. Bernstein, T. E. Smith, R. L. Tobin, B. W. Wie. 1993. A variational inequality formulation of the dynamic network user equilibrium problem. *Operations Research* 41 (1): 179–191.

Fritz, M., and G. Schiefer. 2008. Food chain management for sustainable food system development: a European research agenda. *Agribusiness* 24 (4): 440–452.

Harris, D. R. 1979. Continuities and change in tropical savanna environments. *Current Anthropology* 20 (2): 394–397.

Jewell, C., J. Brown, M. Keeling, L. Green, G. Roberts, K.I.P. Verheyen, C. Fourichon. 2013. Bayesian epidemic risk prediction-knowledge transfer and usability at all levels. In, Society for Veterinary Epidemiology and Preventive Medicine. Proceedings of a meeting held in Madrid, Spain, 20-22 March 2013. *Society for Veterinary Epidemiology and Preventive Medicine* 127–142.

Kao, R. R. 2002. The role of mathematical modelling in the control of the 2001 FMD epidemic in the UK. *Trends in Microbiology* 10 (6): 279–286.

Kao, R. R., D. M. Green, J. Johnson, and I. Z. Kiss. 2007. Disease dynamics over very different time-scales: foot-and-mouth disease and scrapie on the network of livestock movements in the UK. *Journal of The Royal Society Interface* 4 (16): 907–916.

Kim, H. and R. Anderson. 2012. Temporal node centrality in complex networks. *Physical Review E* 85 (2): 026107.

Kleinberg, J. M. 1999. Authoritative sources in a hyperlinked environment. *Journal of the ACM* 46 (5): 604–632.

Kleinberg, J. M. 2000. Navigation in a small world. *Nature* 406 (6798): 845–845.

Lazzarini, S., F. Chaddad, M. Cook. 2001. Integrating supply chain and network analyses: the study of netchains. *Journal on Chain and Network Science* 1(1): 7–22.

Lehmann, R.J., R. Reiche, G. Schiefer. 2012. Future internet and the agri-food sector: State-of-the-art in literature and research. Computers and Electronics in Agriculture 89 (2): 158–174.

Lentz, H. H. K., M. Konschake, K. Teske, M. Kasper, B. Rother, R. Carmanns, B. Petersen, F. J. Conraths, and T. Selhorst. 2011. Trade communities and their spatial patterns in the German pork production network. *Preventive Veterinary Medicine* 98 (2–3): 176–181.

Luning, P. A., K. Kirezieva, G. Hagelaar, J. Rovira, M. Uyttendaele, and L. Jacxsens. 2015. Performance assessment of food safety management systems in animal-based food companies in view of their context characteristics: A European study. *Food Control* 49: 11–22.

MAF. 2009. Review of Selected Cattle Identification and Tracing Systems Worldwide. Lessons for the New Zealand National Animal Identification and Tracing (NAIT) Project. MAF Biosecurity New Zealand Information Paper No: 2009/3. Wellington, New Zealand.

Marsden, Peter V. 2002. Egocentric and Sociocentric Measures of Network Centrality. *Social Networks* 24 (4): 407–422.

Martínez-López, B., A. M. Perez, and J. M. Sánchez-Vizcaíno. 2009. Social Network Analysis. Review of General Concepts and Use in Preventive Veterinary Medicine. *Transboundary Emerging Diseases* 56 (4): 109–120.

Natale, F., A. Giovannini, L. Savini, D. Palma, L. Possenti, G. Fiore, and P. Calistri. 2009. Network analysis of Italian cattle trade patterns and evaluation of risks for potential disease spread. *Preventive Veterinary Medicine* 92 (4): 341–350.

National Institute of Standards and Technology-Big Data Working Group. 2013. http://bigdatawg.nist.gov/home.php. [accessed October 2, 2015]

Newman, M. E. J. 2001. Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. *Physical Review E.* 64 (1): 1–7.

Newman, M. E. J., 2005. A measure of betweenness centrality based on random walks. *Social Networks* 27 (1): 39–54.

Ng, D., Siebert, J. W. 2009. Toward better defining the field of agribusiness management. *International Food and Agribusiness Management Review* 12 (4):123–142.

Nikiforuk, A. 2008. Pandemonium: Bird Flu Mad Cow and Other Biological Plagues of the 21st Century. Penguin, Canada.

Okoth, I. O. and S. Wagner. 2009. Locally Oriented Noncrossing Trees. *The Electronic Journal of Combinatorics* 22 (3): P3.36.

Opsahl, T., F. Agneessens and J. Skvoretz. 2010. Node centrality in weighted networks: Generalizing degree and shortest paths. *Social Networks* 32 (3): 245–251.

Pastor-Satorras, R. and A. Vespignani. 2001. Epidemic spreading in scale–free networks. *Physical Review Letters* 86 (14): 3200–3203.

Petróczi, A., T. Nepusz, G. Taylor, and D. P. Naughton. 2011. Network analysis of the RASFF database: a mycotoxin perspective. *World Mycotoxin Journal* 4(3): 329–338.

Petróczi, A., G. Taylor, T. Nepusz, and D. P. Naughton. 2010. Gate Keepers of EU Food Safety: Four States Lead on Notification Patterns and Effectiveness. *Food and Chemical Toxicology* 48 (7): 1957–1964.

Power, D. J. 2014. Using 'Big Data' for analytics and decision support. *Journal of Decision Systems* 23 (2): 222–228.

Regulation (EC) No 1760/2000 of the European Parliament and of the Council of 17 July 2000 establishing a system for the identification and registration of bovine animals and regarding the labelling of beef and beef products and repealing Council Regulation (EC) No 820/97. *Official Journal of the European Communities* L 204, 11-August 2000.

Regulation (EC) No 882/2004 of the European Parliament and of the Council of 29 April 2004 on official controls performed to ensure the verification of compliance with feed and food law, animal health and animal welfare rules. *Official Journal of the European Union*, L 165, April 30, 2004.

Rommel, S. A. and J. D. McCleave. 1973. Prediction of oceanic electric fields in relation to fish migration. *Journal du Conseil* 35 (1): 27–31.

Salathé, M. and Jones, J. H. 2010. Dynamics and Control of Diseases in Networks with Community Structure. *PLOS Computational Biology* 6 (4): e1000736.

Schroeder, T. C. and G. T. Tonsor. 2012. International Cattle ID and Traceability: Competitive Implications for the US. *Food Policy* 31: 31–40.

Smith, D.A., D.R. White. 1992. Structure and dynamics of the global economy: network analysis of international trade 1965–1980. *Social forces* 70 (4): 857–893.

Sonka, S. 2014. Big Data and the Ag Sector: More than Lots of Numbers. *International Food and Agribusiness Management Review* 17 (1): 1–20.

Stanley, W. and F. Katherine, 1994. *Social Network Analysis: Methods and Applications*. Cambridge University Press, Cambridge, UK.

Sugar, P. F., P. Hanak, and T. Frank 1994. *A History of Hungary.* Indiana University Press, Bloomington, Indiana, USA.

Tichy, N. M., L.T. Michael, and F. Charles. 1979. Social Network Analysis for Organizations. *Academy of Management* 4 (4): 507–520.

Valdano, E., C. Poletto, A. Giovannini, D. Palma, L. Savini, and V. Colizza. 2015. Predicting Epidemic Risk from Past Temporal Contact Data. *PLOS Computational Biology* 11 (3): e1004152.

Valente, T. W., K. Coronges, C. Lakon, and E. Costenbader, 2008. How Correlated Are Network Centrality Measures? *Connect (Tor)* 28 (1): 16–26.

Van Asselt, E. D., P. Sterrenburg, M.Y. Noordam, and H. J. Van der Fels-Klerx. 2012. Overview of available methods for risk based control within the European Union. *Trends in Food Science & Technology* 23 (1): 51–58.

Wagner, C. S., and L. Leydesdorff. 2005. Network structure, self-organization, and the growth of international collaboration in science. *Research Policy* 34 (10): 1608–1618.

Wang, L., S. Jajodia, A. Singhal, P. Cheng, and S. Noel. 2014. k-Zero Day Safety: A Network Security Metric for Measuring the Risk of Unknown Vulnerabilities. *IEEE Transactions on Dependable and Secure Computing* 11 (1): 30–44.

Wang, L., S. Noel, and S. Jajodia. 2006. Minimum-cost network hardening using attack graphs. *Computer Communications* 29 (18): 3812–3824.

Ward, J.S., and A. Barker. 2013. Undefined by data: a survey of big data definitions. arXiv preprint arXiv:1309.5821.

Wasserman, S., and K. Faust. 1994. Social network analysis: methods and applications. Cambridge: Cambridge University Press.

Wen, J.T., and M. Arcak. 2004. A unifying passivity framework for network flow control. *IEEE Transactions on Automatic Control* 49 (2): 162–174.

Wu, X., Zhu, X. Wu, G.-Q. and Ding, W. 2014. Data mining with big data. *IEEE Transactions on Knowledge and Data Engineering* 26 (1): 97–107.

International Food and Agribusiness Management Review
Volume 19 Issue A, 2016

The Role of Wireless Broadband Connectivity on 'Big Data' and the Agricultural Industry in the United States and Australia

Tyler B. Mark[ⓘ a], Terry W. Griffin[b], and Brian E. Whitacre[c]

[a]*Assistant Professor, University of Kentucky, Agricultural Economics,*
417 Charles E. Barnhart Bldg. Lexington, KY 40546-0276, USA

[b]*Assistant Professor, Kansas State University, Agricultural Economics*
342 Waters Hall, Manhattan, KS 66506, USA

[c]*Associate Professor, Oklahoma State University, Agricultural Economics*
504 Ag Hall, Stillwater, OK 74078, USA

Abstract

Big Data has the potential to change the fabric of agriculture as we know it today; but only with wireless connectivity sufficient to employ telematics and other precision agricultural technologies. The primary focus of this paper is on data transfer, specifically as an enabling technology to precision agriculture. Limited wireless internet connectivity impedes the full utilization and effectiveness of precision agricultural practices and subsequent agricultural big data system. These failures lead to potential differentiation of farmland values; those fields with adequate wireless connectivity commanding a premium. In the absence of wireless data transfer for download and upload, precision agriculture technologies such as telematics cannot be utilized efficiently. Improving wireless connectivity is a primary driver of the adoption of big data. Increased connectivity could also intensify the adoption of precision agricultural technologies leading to input cost savings and decreased input usage.

Keywords: wireless, internet, rural broadband, telematics, precision, connectivity

ⓘCorresponding author: Tel: + 1 859.257.7283
Email: T. B. Mark: Tyler.Mark@uky.edu
T. W. Griffin: twgriffin@ksu.edu
B. Whitacre: brian.whitacre@okstate.edu

Introduction

The United States and Australia are global leaders in food production and agricultural technology adoption. For the 2014/2015 crop years the United States and Australia combined to produce 19%, 22%, and 15% of the world's coarse grains, oilseeds, and cotton, respectively (USDA/FAS 2016). These countries are also significant adopters of precision agricultural technologies, especially Global Navigation Satellite System (GNSS but formerly referred to as Global Positioning Systems or GPS) enabled yield monitors and automated section control, to name a few, are now standard on new equipment. This is a result of the continuous innovation that takes place in the agricultural sector. Humans have advanced from being hunters and gatherers to the point where we are annually increasing the yields of primary commodities, such as corn, soybeans, wheat, cotton, etc. via improved genetics and production practices (Fischer and Edmeades 2010). The increases seen in crop yields is a function of technology adoption starting with improved management of these crops, the adoption of precision agriculture technologies, and genetic modification (Evenson and Gollin 2003). To keep up with population growth it is expected that food production must double between 2014 and 2050 (FAO 2009). To accomplish this, production agriculture will have to continue efficiency improvements with respect to practices and inputs in order to optimize output per acre and limit the negative externalities created through production intensification. Success will depend on efficiently and effectively converting vast amounts of data that are generated into information and subsequently knowledge to make real-time decisions and justify the utilization of inputs.

Big Data has the potential to change the fabric of agriculture as we know it today. *Big Data* is data whose size, scale, and unstructured nature require the usage of new analytical tools and frameworks to be developed and employed (Sonka 2014). These frameworks need to be flexible enough to weave together data from millions of acres and from various sources, such as weather data, yield data, satellite imagery, small unmanned aerial systems (sUAS) imagery, planting prescriptions, and equipment diagnostics just to name a few. If this can be done, it has the potential to be the next agricultural revolution of smart products utilizing precision application of inputs, yield monitors, and other site-specific sensors, analogous to how smartphones changed popular culture, to a point where the system can be evaluated as a whole. However, this revolution requires a new or at least different mindset compared to the way most agricultural producers are operating today (Griffin et al. 2016). Many producers today are rightfully only focusing on their own operation and have reservations regarding several key questions including but not limited to:

- Data ownership: producer vs manufacturer vs landowner vs retailer;
- Data utilization, privacy, storage, and security: data access, utilization, sharing;
- Data value;
- Data transfer.

Each of these stated issues are substantial barriers to the technology (Ferrell 2016; Stubbs 2016); however, the primary focus of this paper is on the last topic, data transfer, specifically as an enabling technology to precision agriculture. This paper proceeds assuming data ownership, utilization, valuation, and especially data privacy have been satiated at least in the short run. Limited wireless internet connectivity impedes the full utilization and effectiveness of precision agricultural practices and the subsequent agricultural big data systems. In simplest terms, wireless connectivity is an enabling technology for precision agriculture; and lends itself to be the next infrastructure that may limit the usefulness of agricultural technology. In the absence of wireless data transfer for download and upload, precision agriculture technologies such as telematics cannot be fully utilized. Whitacre et al. (2014) borrow their definition of telematics from Heacox (2008) as "transmitting of data through wireless communication links between the home base and field units", however, an industry standard definition is likely to be offered by AgGateway as "the transmission and receiving of data over long distance communication links" (AgGateway 2016).[1] Wireless transmission of agricultural data, i.e. telematics, is seen as a necessary condition for the maturation of big data capabilities in agriculture. Telematics not only requires adequate wireless internet connectivity bandwidth, but this connectivity is required in non-residential areas where cellular services have not been offered at the same performance level as urban areas (Whitacre et al. 2014). Without sufficient internet connectivity, the transfer of agricultural data remains possible although with additional caveats.

According to Erickson and Widmar (2015), one of the most notable changes over the last three surveys of agricultural service providers is the usage of telematics for field-to-home office communications. In 2011, only 7% of service providers offered telematics data services but the percentage increased to 15% by 2013, and to 20% in 2015. There were slightly more dealerships offering telematics in the Midwestern US (17%) than in other states (12%) in 2013, potentially due to the lack of broadband connectivity outside the Midwest (Whitacre et al. 2014). In addition, the common wireless carriers utilized by the leading equipment manufacturers have a larger presence in rural areas of the Midwestern US than they do in other regions. Holland et al. (2013) reported that two-thirds of service providers stated telematics are perceived to be an emerging technology with 30% suggesting an uncertain future and 37% suggesting a promising future; indicating uncertainty with respect to the future of the technology among service providers (Whitacre et al. 2014). Until wireless internet is sufficient to transfer agricultural data, the impedance of telematics and precision agriculture are likely to capitalize into substantial farmland value differences for internet-connected and internet-deficit fields (Griffin et al. 2016). Griffin et al. (2016) describe scenarios where the absence of biophysical and geo-spatial site-specific data could result in penalties during farmland sales or rental auctions. They also describe how farmers are not expected to pay similar rental rates for farmland without adequate wireless connectivity *ceteris paribus*. In part, it is the wireless connectivity that empowers farmers to securely archive the biophysical geospatial data in the former example.

The realization of the *Big Data's* full value will not happen until the wireless connectivity barrier is overcome. Expanding upon Whitacre et al. (2014) we have two objectives. First, given the new broadband definition for the United States, we explore the wireless coverage for the United States and Australia similar to Whitacre et al. (2014). We focus on examining broadband availability for crop production regions. These high-production broad acre areas are also the areas where precision agriculture adoption rates are expected to be the highest and where telematics most likely to be employed; therefore where the value of big data is initially expected to be fully realized. It should also be noted that some high value crop utilize various aspects of precision agriculture. For example, viticulture is a classic example of a high value crop that utilizes precision agriculture, especially yield monitors, irrigation, and other monitoring sensors (Bramley and Proffitt 1999; Bramley 2001; Bramley et al. 2003, 2005). However, without adequate access to wireless internet, the development of a big data system will lag behind potential development. Additionally, insufficient connectivity could

[1] AgGateway AgGlossary. Telemetry - The transmission and receiving of data over long distance communication links. http://agglossary.org.

limit the value of a big data system if the technologies used to populate the system are inefficient in data transfer.

Status of Broadband Connectivity in US and Australia

Broadband connectivity across the globe but specifically within the United States and Australia will have a significant impact on both big data utilization and on the agricultural industry at large. In the absence of broadband connectivity and wireless data transfer, the benefits of big data and telematics services are limited. In general the industry is experiencing effects related to network externalities. In addition to constraining the profitability of agricultural firms; lack of broadband connectivity limits the adoption and efficiency of precision agricultural technologies that make use of or rely upon near real time connectivity. Additionally, these precision agriculture technologies are the primary data collection methods populating this big data system.

In the United States, the National Broadband Map (NBM) provides data on wireless availability over a range of broadband speeds. Superimposing these data on top of publicly available crop production data from the United States Department of Agriculture (USDA) illustrates the need for increased wireless connectivity in nonresidential areas. Using analogous data from Australia including Australian Government Department of Communications and Australian Bureau of Statistics we compare and contrast these two nations known to be leaders in production agriculture and technology utilization.

Current Status of Broadband in US

The United States Federal Communications Commission (FCC) updated the definition of broadband in January 2015. The faster speeds required to be considered broadband brought light to connectivity barriers, especially with respect to broadband connectivity gaps in specific geographic areas such as agricultural production regions. Specifically, the 25 megabit per sec (Mbps) download speed requirement negates the majority of United States wireless connections from being classified as broadband. Figure 1 shows the discrepancy between the download and upload speeds required by the FCC to be considered broadband.

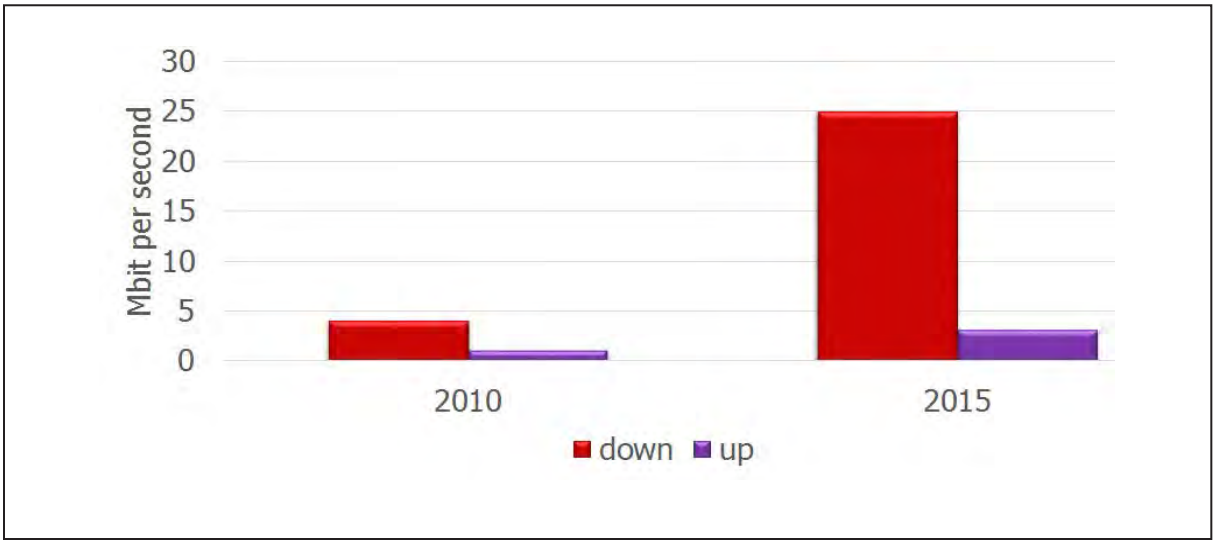


Figure 1. US FCC-defined broadband speeds

Recently passed state-level legislation, such as Iowa's "Connect Every Acre" bill that was signed into law in June 2015, demonstrates the recognition of this topic by today's policy makers. In addition, recent congressional hearings on internet connectivity in general and another specifically on big data in agriculture both discussed the ramifications of internet on agriculture (see Ferrell (2015), testimony to U.S. House of Representatives). Many producers currently employing precision agriculture technologies do not have access to broadband speed wireless internet. Figure 2 shows the relationship between corn and wheat production and download speeds. The dark black blocks represent areas where greater than 75% of the population meets the 25 mbps download requirements and the majority of these blocks are located in close proximity to major cities and not prime agricultural areas.

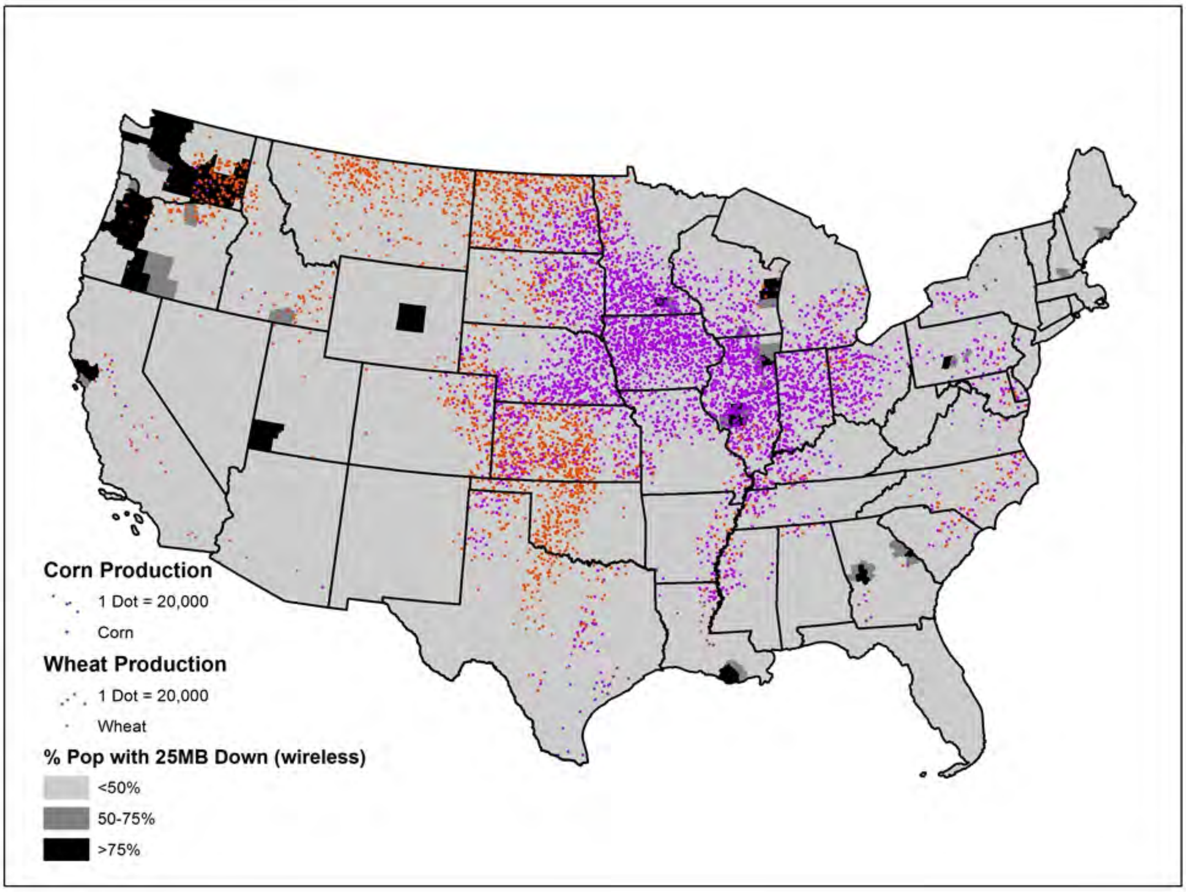


Figure 2. Wireless download availability for corn and wheat production, 2015.

Figure 3 shows the relationship between corn and wheat production and upload speed broad-band speed requirements. The low hurdle of 3 mbps results in significant parts of the country achieving the hurdle to be considered broadband. Shearer (2014) points out that most precision agriculture data needs to be uploaded rather than downloaded; and given that upload speeds are substantially slower than download speeds, moving data such that real-time decisions can be problematic. For some types of data such as machine diagnostics, planting prescriptions, and the like the current speeds offered are probably adequate. However, yield data and specifically imagery data may require connectivity speeds in excess of what the industry currently offers. More importantly, these connectivity requirements may not be a cost effective method of data transfer, given labor and connectivity costs.

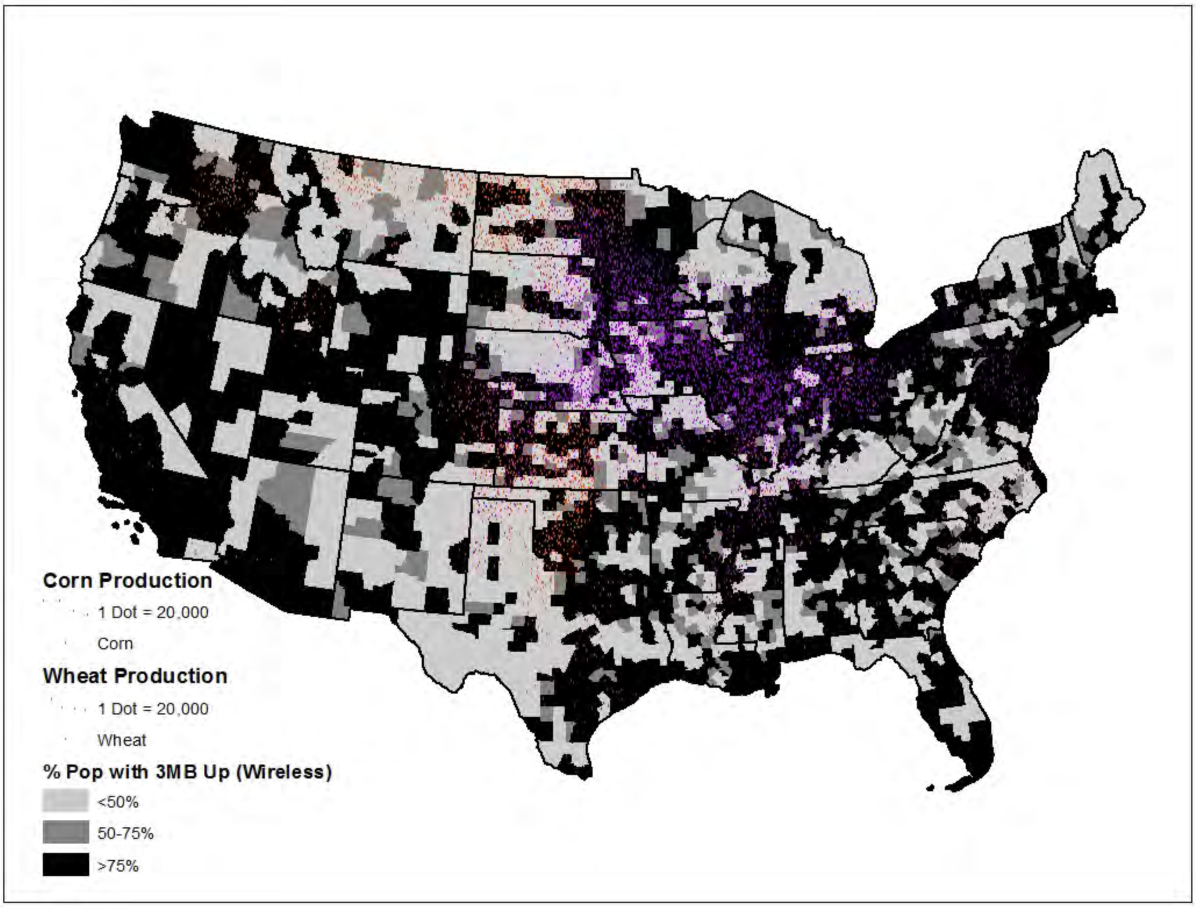


Figure 3. Wireless upload availability for corn and wheat production, 2015.

Current Status of Broadband in Australia

Similar to the United States, there are multiple mobile network providers in Australia. Existing coverage providers include Telstra, Optus, Vodafone, and National Broadband Network (NBN). The priority of these providers was initially voice coverage rather than data coverage especially in rural crop producing areas, just as in the United States. As an example Telstra (the largest Australian coverage provider) coverage areas is mapped in Figure 4. As expected, the cellular coverage area mirrors that of the residential population areas. A primary difference between the United States and Australia is that the Australian government provides internet infrastructure for resale by retailers. Australia's NBN is a government owned wholesale provider of high speed coverage that sells via retail service providers. The NBN provides coverage through fiber optics, fixed wireless, and satellite.[2]

[2] see http://www.nbnco.com.au/ for more detail on NBN

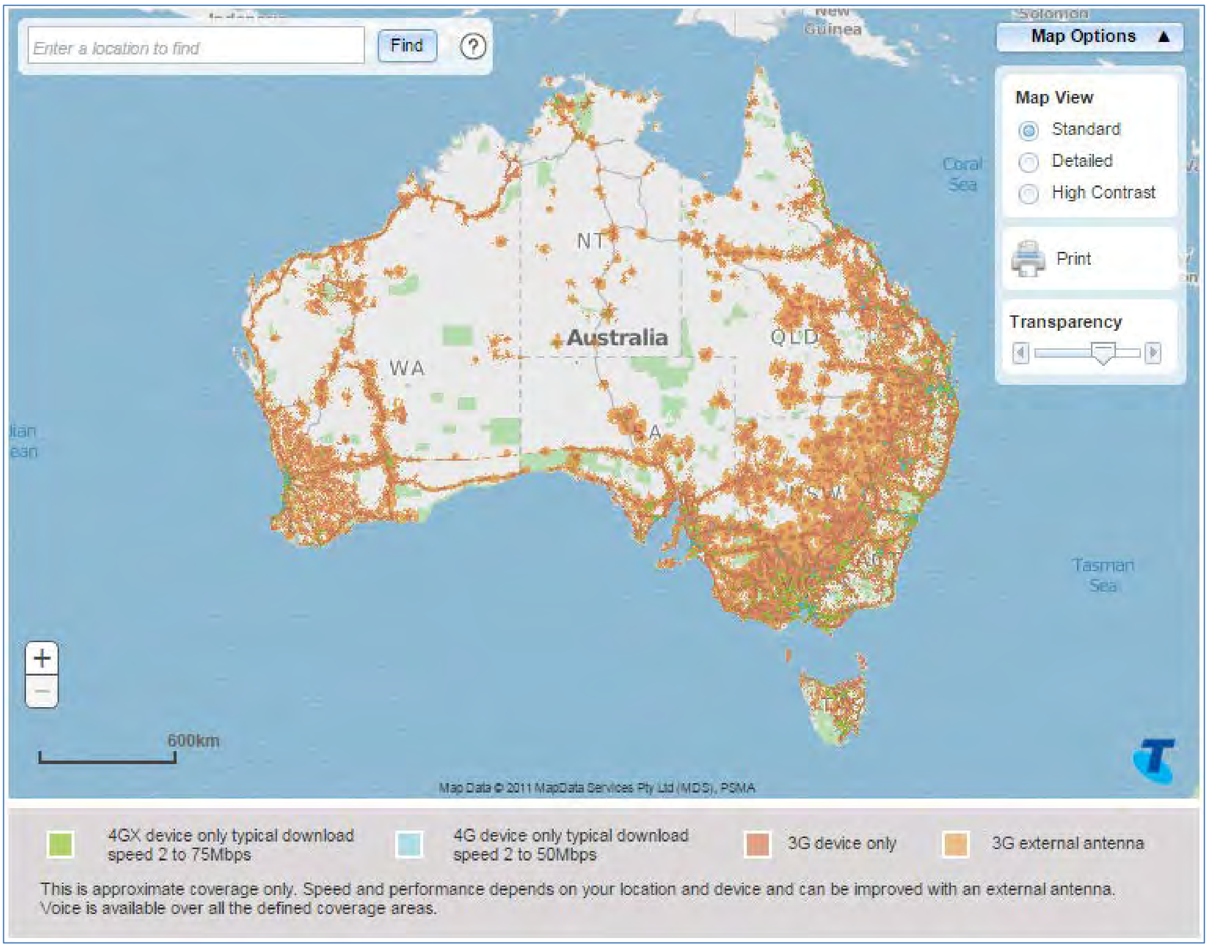


Figure 4. Wireless cellular coverage area from Telstra (the largest coverage provider in Australia).

Source. https://www.telstra.com.au/coverage-networks/our-coverage

The land use across Australia is presented in Figure 5. Similar to the United States, there are areas that are neither populated or cultivated. The areas of interest are the cultivated areas that have low population and therefore minimal if any wireless internet connectivity. For instance, the areas signified by the yellow, orange, and gold colors are agricultural production areas where *Big Data* are most likely to be adopted due to precision agricultural practices. The red and grey areas indicate urban and rural residential areas, respectively, to emphasize the relative location of populous and agricultural production.

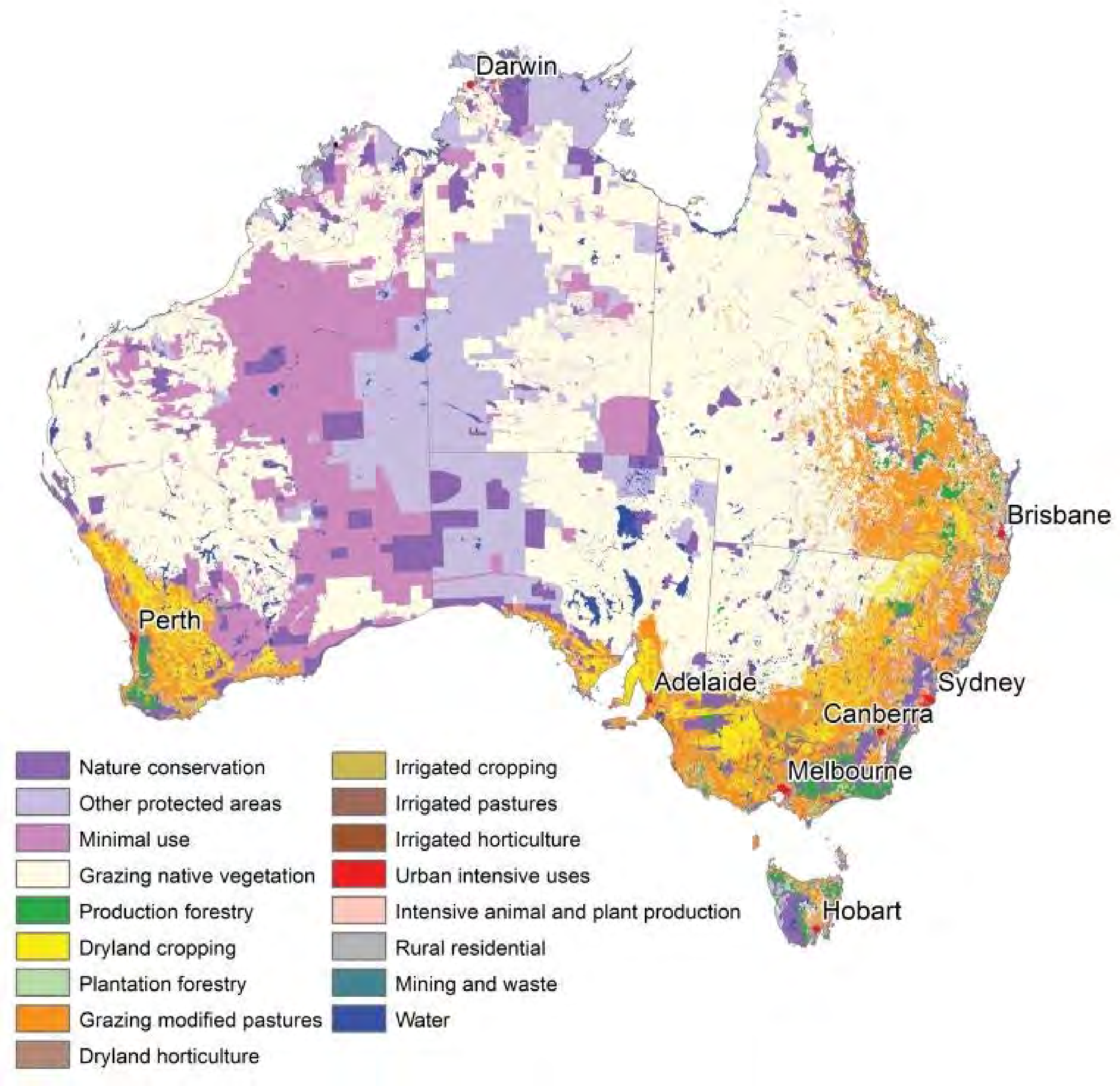


Figure 5. Land use in Australia, 2005-2006.

Source. Australian Government Department of Foreign Affairs and Trade

Rest of World and Global Initiatives

The US and Australia are not unique across the globe in providing high speed internet. In Germany, the government is pushing a 50 Mbps target up from the existing threshold of 11 Mbps for all users (Woods 2015). This doubles the currently highest global average speed of South Korea of 24 Mbps (Woods 2015). Wireless internet providers have only started to provide wireless connectivity sufficient to move data from crop producing regions far from residential areas. These providers have been encouraged and supported by the efforts of Google and Facebook to connect the world's population. Facebook has been reported to evaluate drones (Lee 2015), satellites, and high-altitude balloons (Patterson 2015) for providing internet access in developing regions. Google has similar goals; to make internet connectivity ubiquitous for every global citizen.

To this point this study has focused on two heavily developed countries. However, this is not to say that information gleaned from the adoption of telematics and big data analytics cannot be transferred to developing countries. First, most likely the types of technologies that Facebook and Google are developing will be needed for developing countries who lack the capital to build the infrastructure needed for increased connection speeds. This could even be said for some part of the United States. In the meantime, as we start to populate and gain a deeper understanding of the agricultural systems both farmers in developing countries and small farms in developed countries will benefit. Historically, precision agricultural technologies have had economies of scale barriers to entry. With the introduction of big data farmers in developing countries and small farmers will have the opportunities to benefit from the findings. We suspect that the gains from farmers in developing countries and small farmers will be relatively greater than those who are already utilizing precision agricultural technologies. This would simply be a function of now being able to make better management decisions with the newly acquired information.

Data Transmission Needs

Current forms of data transmission (i.e. cellular, wireless, and satellite) are lagging behind the needs of production agriculture (Griffin and Mark 2014). CoBank (2016) just released an infrastructure briefing where they interviewed producers about their data usage and needs. The found that during peak harvest large producers can utilize 30 plus gigabytes of data per month. This is more than double what they were using three years ago and the need is only growing. There has been a significant push to increase availability of broadband internet connectivity in rural areas, where the majority of agricultural production takes place. This has been a very slow process that is not keeping pace with demand to the point that connectivity is a barrier to the full utilization of current precision agriculture technologies or at least the internet is seen as an enabling technology. Cellular and hard-wired providers have not been incentivized to expand their services in rural areas due to the lack of sufficient voice service customers needed to justify the investment. However, satellite based internet connectivity could provide an effective solution over larger geographic areas. However, one of the downsides to satellites is the signal can be interrupted and that can cause issues with applications, such as John Deere Machine Sync, that require uninterrupted connectivity to function properly (John Deere 2015)

The typical setup for producers today involves utilizing cellular connectivity to transfer data or the *status quo* of manually transferring data. Current 4G cellular connections only allow up to a 10 Mbps download speed, and upload speeds that range from 2 to 5 Mbps. Historically, differences between upload and download speeds were several magnitudes different due to residential uses relied more on download than upload. This is evident if one considers watching streaming video via Netflix, i.e. downloaded data; however more recent phenomena such as uploading ‘selfies’ to Facebook require relatively more upload. This has increased wireless providers desire to improve upload speeds. Anecdotal evidence suggests wireless connectivity has a 2:1 ratio of download to upload speeds in the US (Speedtest 2016). Increasing the upload speed would provide the agriculture sector with a much needed boost in capacity.

Little information actually exists or is not publically available on file size by type. Additionally, precision agriculture provider has proprietary software used to package and move the data. However, Shearer (2014) estimated that row crop producers potentially generate 0.5 kilobytes of data per plant. In other words, a corn producer with a plant population of 30,000 seeds per acre could produce 15 megabytes of data per acre each year; and if this were a 1,000 acre corn farm, they could potentially produce 15 gigabytes of data per year that would need to be transferred. Extrapolating this out to the approximately 88.9 million acres of corn planted in 2015 there would be approximately 1,333.5 terabytes of data produced. This estimate does not include the usage of drone or UAV imagery data that is increasing in popularity in agriculture; the reliance on imagery from drones greatly increase these data transfer requirements. The amount of data generated from drones will depend on the type, frequency, and quality of images that are being taken (Buschermohle 2014). Recently commercialized technology allows multiple automated section control (ASC) enabled vehicles to share a coverage map so that each vehicle is instructed to turn application on or off to prevent over application or gaps. Bennett (2016) reports economic implications of shared coverage map technology and how satellite ping rates are not sufficient; suggesting that cellular connectivity is the only viable option for coverage map sharing.

A key for most producers will be deciding which data layers will require real-time transfer and analysis. Data layers that might require real-time transfer are yield data and equipment diagnostics. However, some data layers could be wirelessly transferred after the fact once a connection is achieved. Furthermore, without sufficient wireless data transfer service, producers rely on manual data transfer which may not happen until after the season is over and furthermore cause suspicion with third parties. Many third-parties require data in real time from the sensor-based equipment to prevent data from being corrupted either intentionally or unintentionally. By the time data are manually moved to the analytics, opportunities to adjust management practices are missed, significantly affecting farm profitability, productivity, and environmental impact. In addition, real-time communication between farm equipment and online servers is not possible. Finally, land that lacks adequate connectivity leads to geospatial data not being sufficiently backed-up in a timely manner, therefore increasing the risk of this valuable data being lost, destroyed, or otherwise not used.

Discussion

These findings suggest that opportunities exist for the private and/or public sectors to increase wireless connectivity infrastructure. This could be in the form of improved satellites, increased wireless cellular infrastructure, or high altitude balloons. The primary criteria for their usage will be upload capacity and reliability. Improving wireless connectivity could be one of the primary drivers of the adoption of big data, or at least not to impede adoption. The increase in connectivity could also increase the adoption of precision agricultural technologies that can lead to input cost savings and decreased input usage. Without adequate connectivity to allow efficient and cost effective data transfer, the value of the big data system will be limited for both direct and indirect users, such as producers and consumers, respectively. In the U.S., internet service providers, especially wireless providers, are all private sector firms as opposed to Australia where the federal government provides basic internet infrastructure for resale.

These results are of interest to public policy makers, environmental groups, private sector satellite internet service providers, and members of the agricultural industry including farmers, equipment manufacturers, and software companies. Quantifying the magnitude of the problem and providing guidance toward a feasible solution will aide in maintaining a sustainable production agriculture industry now and for years to come.

Moving Forward

It is expected that over the next five years there will be an increased desire and demand for producers to collect and analyze data in an effort to increase the efficiency of their operation. This will be even more important during time periods of low commodity prices and increased scrutiny over chemical and nutrient usage. If producers are able to track and verify usage of these inputs, it could help minimize perceived environmental impacts and legal costs for producers. Wireless data transfer technologies including satellite, cellular, hard-wired, and potential balloon based systems are suitable candidates to fill the connectivity void and all need to be explored. All of these technologies have the potential to be used for agricultural data transfer but none are universally perfect for the task. However, specific characteristics are sought to move forward with the increased data collection requirements within the agriculture sector. The first and most important characteristic is reliable access. As seen in Figure 2 and Figure 3 there are significant gaps in current broadband offerings and most of these voids exist in prime agricultural areas. Currently, upload and download speeds are a significant bottleneck for farmers depending on the file type and size, in terms of receiving real-time feedback. This does not mean that internet connectivity is required for planting or harvest. However, if the upload or download speeds are too slow, field efficiency can be decreased and in turn decrease the number of acres covered in a given day. This can be especially true in fields that are running multiple units and are requiring them to communicate in real time where the other unit has been in the field. During planting season slow internet speeds can mean the difference in getting the crop in before a weather event or suffering yield penalties for an untimely planting, thus translating into potentially decreased income for the operation. There have been substantial pushes by state governments to increase broadband access to help increase business development in rural areas. However, the expansion of broadband access has been very slow and is not keeping pace with the demands of the industry.

References

Australian Government Department of Foreign Affairs and Trade. Land Use in Australia, 2005–2006. https://www.telstra.com.au/broadband/nbn/nbn-rollout [accessed Feb. 2016].

Bennett, J.D. 2016. Value Of Map Sharing Between Multiple Vehicles In The Same Field While Using Automated Section Control. MAB Thesis, Manhattan, KS, Kansas State University.

Bramley, R.G.V. and A.P.B. Proffitt. 1999. Managing variability in viticulture production. *Grapegrower and Winemaker* 427: 11–16.

Bramley, R.G.V. 2001. Progress in the development of precision viticulture – variation in yield, quality and soil properties in contrasting Australian vineyards. In *Precision Tools for Improving Land Management,* edited by L.D. Currie and P. Loganathan, 25-43. Occasional report No. 14. Fertilizer and Lime Research Centre, Massey University, Palmerston North.

Bramley, R., B. Pearse and P. Chamberlain. 2003. Being profitable precisely – a case study of precision viticulture from Margaret River. *Australian Grapegrower and Winemaker.* (473a):84–87.

Bramley, R.G.V., A.P.B. Proffitt, C.J. Hinze, B. Pearse, and R.P. Hamilton. 2005. Generating benefits from precision viticulture through selective harvesting. In *Proceedings of the 5th European Conference on Precision Agriculture,* edited by J.V. Stafford, 891–898. Wageningen Academic Publishers, The Netherlands.

Buschermohle, M. 2014. Enhancing the Value of Precision Ag Data with Unmanned Aerial Systems. Paper presented at 2014 Crop Management Seminar. November. http://www.plantmanagementnetwork.org/edcenter/seminars/2014CottonIncCropMgtSeminar/.

CoBANK. 2016. Farmers Harvest Gigabytes with Broadband and Wireless Technologies. *Rural Infrastructure Briefings.* March. http://www.cobank.com/Knowledge-Exchange/Research-Reports.aspx [accessed March 17, 2016].

Erickson, B., and D.A. Widmar. 2015. Precision Agricultural Services Dealership Survey Results. https://www.agecon.purdue.edu/cab/ArticlesDatabase/articles/2013PrecisionAgSurvey.pdf [accessed May 25, 2016].

Evenson, R.E. and D. Gollin. 2003. Assessing the impact of the green revolution, 1960-2000. *Science.* 300:758-762. DOI: 10.1126/science.1078710.

Food and Agricultural Organization (FAO). 2009. FAO's Director-General on how to feed the world in 2050. *Population and Development Review.* 4(35):837-839.

Ferrell, S.L. 2015. Testimony to U.S. House of Representatives Big Data and Agriculture: Innovations and Implications. http://agriculture.house.gov/uploadedfiles/10.28.15_ferrell_testimony.pdf [accessed May 25, 2016].

Ferrell, S.L. 2016. Forthcoming. Legal Issues on the Farm Data Frontier, Part I: Managing First-Degree Relationships in Farm Data Transfers. *Drake Journal of Agricultural Law*. 20(1): Spring.

Fischer, R.A. and G.O. Edmeades. 2010. Breeding and Cereal Yield Progress. *Crop Science* 50:S85-S98. DOI: 10.2135/cropsci2009.10.0564.

Griffin, T.W. and T.B. Mark. 2014. Value of Connectivity in Rural Areas: Case of Precision Agriculture Data. Paper presented at International Conference on Precision Agriculture. July. Sacramento, CA.

Griffin, T.W., T.B. Mark, S. Ferrell, T. Janzen, G. Ibendahl, J.D. Bennett, and J.L. Maurer. 2016. Big Data Considerations for Rural Property Professionals. *Journal of the American Society of Farm Managers and Rural Appraisers*. 167-180

Heacox, L. 2008. Time for Telematics. *CropLife*. Sept. 10. http://www.croplife.com/equipment/precision-ag/time-for-telematics/.

Holland, J.K., Erickson, B., and D.A. Widmar. 2013. Precision Agricultural Services Dealership Survey Results. http://agribusiness.purdue.edu/files/resources/rs-11-2013-holland-erickson-widmar-d-croplife.pdf [accessed May 25, 2016]

John Deere. 2015. Satellite Module Theory of Operation. http://manuals.deere.com/omview/ OMPFP14775_19/AE77568,0000024_19_20150303.html [accessed March 9, 2016]

Kelly, H. 2014. Facebook looks to drones, lasers and satellites for Internet access. http://www.cnn.com/2014/03/27/tech/innovation/facebook-drones-lasers-connectivity/ March 28, 2014.

Lee, D. Facebook builds drone for internet access. http://www.bbc.com/news/technology-33728704. [accessed May 25, 2016]

Patterson, Thom. 2015. Google, Facebook, SpaceX, One Web plan to beam Internet everywhere. http://www.cnn.com/2015/10/30/tech/pioneers-google-facebook-spacex-oneweb-satellite-drone-balloon-internet/ [accessed May 25, 2016].

Shearer, S.A. 2014. Big Data: The Future of Precision Agriculture. Presented at The InfoAg Conference. St. Louis, MO, June. http://past.infoag.org/abstract_papers/papers/paper_233.pdf

Sonka, S. 2014. Big Data and the Ag Sector: More than Lots of Numbers. *International Food and Agribusiness Management Review* 17(1): 1–20.

Speedtest. 2016. Americas Fastest ISPs and Mobile Networks 2015. http://www.speedtest.net/awards/us. [accessed May 26, 2016].

Stubbs, M. 2016. Big Data in U.S. Agriculture. Congressional Research Service. 7–5700 R44331. http://www.fas.org/sgp/crs/misc/R44331.pdf [accessed February 28, 2016]

U.S. Department of Agriculture, Foreign Agricultural Service. 2016. *Production, Supply and Distribution.* USDA/FAS. Reports: Field Crop-Production.

Whitacre, B.E., T.B. Mark, and T.W. Griffin. 2014. How Connected are Our Farms? *Choices* 29(3). www.choicesmagazine.org/UserFiles/file/cmsarticle_392.pdf.

Woods, B. 2015. Germany promises 50Mbps broadband for all, 10 times faster than global average speeds. http://thenextweb.com/eu/2015/09/08/germany-promises-50mbps-broadband-for-all-10-times-faster-than-global-average-speeds/. [accessed October 15, 2015].

International Food and Agribusiness Management Review
Special Issue - Volume 19 Issue A, 2016

Big Data in Agriculture: Property Rights, Privacy and Competition in Ag Data Services

Michael E. Sykuta[ⓘ]

Associate Professor, Agricultural & Applied Economics Department and Director of the Contracting and Organizations Research Institute, University of Missouri, 135 Mumford Hall, Columbia, MO, 65211, USA

Abstract

The transition from precision agriculture to *big data* opens a variety of concerns among both farmers and ag data service providers about the privacy, ownership and use of farm data. This paper discusses the nature of those concerns as well as an attempt to establish industry guidelines concerning data privacy and security. I argue that the proposed guidelines likely do little to protect any real privacy concerns and run the risk of negative consequences for competition in ag data services and of reducing valuable information flows in commodity markets.

Keywords: big data, ag data services, property rights, competition

[ⓘ]Corresponding author: Tel: + 1 573.882.1738
Email: M. E. Sykuta: sykutam@missouri.edu

Introduction

While some may picture farming as a bucolic enterprise, far removed from big data and the information economy, modern farming practices are intensely data-driven. Precision agriculture practices have implemented "embodied-knowledge" technologies such as GPS-guided equipment and variable-rate planting and spraying equipment, while "information intensive" technologies such as on-board yield monitors and grid soil sampling generate an enormous amount of site-specific data that can be analyzed to assess input needs and output performance (Griffin et al. 2004). According to the USDA's 2012 Agricultural Resource Management Survey, over 62% of corn and soybean acres in the US were harvested with yield monitoring devices and 73% of acres were farmed using some type of precision agriculture practice.

However, despite the enormous amount of data already generated across tens of thousands of farms each year in the US, the promise of big data remains largely unfulfilled. While the volume, velocity, and variety of data generated in production agriculture have been available for years, the ability to aggregate, analyze, and distill value-creating decision support tools from that data is still in the early stages. Much of the data generated over the past decade sits on the computers of individual farmers who have little ability to make use of it independently. Whitacre et al. (2014, 3), argue that "[b]efore 'big data' will be widely accepted by farmers and others across the agricultural industry, its collection, processing, on-demand analytics, and decision making must become passive to the user." In other words, until a new sector of data service providers is able to harness and commercialize the revenue-generating correlations to be discovered in these fields of data, the full productivity-enhancing potential of big data will remain but a promise.

But it's a promise worth a lot of potential value and several traditional agricultural input providers have been working to develop data services with the hope of capturing a share of that promised value. John Deere, one the industry leaders in providing data-generating technologies on their farm machine equipment, and Monsanto and DuPont Pioneer, which together control the majority of the US corn and soybean seed markets, are competing to provide farm-specific decision support tools using big data analytics. The industry has also attracted smaller, start-up enterprises seeking to exploit specific niches of data-driven decision making tools. In announcing its 2013 acquisition of The Climate Corporation, an atmospheric data science company developing micro-weather forecasts to aid farmers' management decisions, Monsanto estimated that the data science market in agriculture could be worth as much as $20 billion.

As more agricultural technology providers (ATPs) enter the market and as there is more focus on aggregating famers' data, farmers have raised concerns about data ownership and privacy.[1] Who owns the data? Who is entitled to the value of the data? How will that data be used or potentially shared? These concerns are particularly strong when dealing with large players such as DuPont Pioneer and Monsanto who have an interest in selling their own agronomic products in addition to the data services themselves; will ATPs engage in

[1] The term "agricultural technology provider" or ATP generally refers to a company that aggregates famer's data, combines it with other relevant data sets, and applies algorithms to analyze the data. Farm input providers like John Deere, Monsanto (Climate Corporation), and Pioneer that provide data-driven agricultural support services would qualify as ATPs due to their use of big data to supplement their farm input businesses.

discriminatory behavior either in pricing or product recommendations based on their knowledge of the local farm operations?

A more general concern is whether these large companies have an unfair competitive advantage, or an ability to foreclose the market for data services, by collecting and controlling data for large percentages of US farm operations. An even greater concern among some agricultural producers is whether the continued development of automated agricultural equipment, driven by big data analytics, may fundamentally change the organization and management of production agriculture.

Some of these concerns were addressed, at least in principle, in an industry-negotiated set of guidelines announced November 13, 2014. These “Privacy and Security Principles for Farm Data” were negotiated by national farmer organizations such as American Farm Bureau Federation, National Farmers Union, and the national trade groups for soybean, corn, wheat, and rice growers; and by several leading ag data companies including John Deere, Monsanto’s Climate Corporation, DuPont Pioneer, and Dow AgroSciences. These principles outline expectations with regard to ownership, portability, disclosure, use/sale, and retention of data as well as contracting practices. It’s unclear how these provisions will work in practice. It’s also unclear whether, or how, the practice of these principles will address concerns about the value of famers’ data and the ability of data service providers to gain monopolistic advantages based on their data repositories. Moreover, some of these principles may run counter to ATPs’ proprietary interests in the information services they provide.

This paper attempts to examine the nature of concerns about big data applications in agriculture and their implications for the market for agriculture data services. The paper will explore the nature of “ownership” of data in the context of information-based products, and how property rights between farm data owners and data service providers may be in conflict. The paper will then explore the potential for large, incumbent ATPs to monopolize or foreclose the market for data services and how proposed data ownership principles may affect that ability. In the process, the paper will address alternate organizational mechanisms that may mitigate any anticompetitive potential.

The Shift to Big Data in Ag

Before addressing data ownership, it’s important to understand the changes in data use that have given rise to the concerns about data ownership and how those new data-driven technologies differ from the precision agriculture practices of the last two decades. Precision agriculture, as the name implies, allows farmers to more precisely manage their crop production, typically in field crops such as corn, soybeans, wheat, and rice. Using yield monitor data, GPS-based field maps and soil sampling, and/or GPS-tagged weed scouting data, farm machinery equipped with variable-rate technologies (VRT) can adjust seed planting density and application rates for herbicides, pesticides and nutrients based on variations in soil quality, topography, moisture, weeds, pests, etc., within field plots. By more precisely planting and managing crops, farmers can increase yields and lower input costs. GPS-assisted or controlled navigation of planting, spraying and harvesting equipment also helps reduce fuel costs by minimizing passes across a field, minimizing overlap of applications, and maximizing the area planted and harvested, further reducing input costs per acre.

Schimmelpfennig and Ebel (2011) report that adopters of yield monitor technologies, GPS mapping, and VRT fertilizer technologies for corn and soybean production produced significantly higher yields than non-adopters in 2001 and 2005, on the order of 10% to 14% higher. On corn land yielding 150 bushels per acre that suggests a revenue increase of $95 to $130 per acre at a corn price of $6.50. Considering that commercial crop operations range from 500 to 5,000 or even 10,000 acres, the value potential is substantial. They also report that adopters of GPS-mapping and VRT fertilizers had significantly lower fuel costs than non-adopters.

Schimmelpfennig and Ebel (2011, 20) suggest that precision agriculture was still in the early stages of adoption. They argue that while profitability of precision ag is likely to affect adoption rates, they point to research by Fernandez-Cornejo, et al. (2001) finding that college educated producers are 15% more likely to adopt precision ag practices. Griffen et al. (2004), argue that precision ag is “human capital intensive”, which would suggest older farmers are less likely to make the investment given their shorter expected time horizon.

However, while precision agriculture practices began taking off in the early 1990s, the data themselves have traditionally stayed close to the farm. Yield monitor data and fields maps are typically generated by the farmer or by a local customized service provider who delivers the data to the farmer. VRT fertilizers may use data generated by soil sensors as the equipment is crossing the field. VRT herbicide and pesticide applicators may use GPS-coded field scouting reports generated either by the farmer herself or a contracted scout who provides the data to the farmer. The data were not aggregated with data from other farms to develop new products, services, or management analytics. In any of these arrangements, there is little doubt about the ownership of the data and little concern about data privacy since data are stored on local computers.

What’s new with big data versus traditional precision agriculture is the aggregation of data from large numbers of farms, the timeliness with which it is aggregated, a surrender of physical custody of the data from the farmer to the ATP—often using cloud technologies—and the potential use of that data for purposes beyond the farm itself.

As the market for ag data services is still in its infancy, there are different pricing models and types of services available from different vendors. In general, the farmer works with a certified local input dealer for a specific input supplier ATP. The farmer generates (or hires the dealer or a third-party company to generate) data on field-specific attributes such as GPS-coded soil sampling and field maps for selected plots of land. The data requirements for ATPs vary based on their algorithms and services offered. The nature of data security issues also differ by vendor given their services and platforms.

For instance, Monsanto’s FieldScripts® program requires two years of raw yield data in addition to soil and field mapping data to generate its planting prescriptions. The farmer also provides information on anticipated planting dates, yield goals, row spacing, and variable-rate planting ranges. Once the data are sent from the local certified dealer to Monsanto, a primary and secondary planting recommendation is developed offering two DEKALB® seed types and planting densities. A preview of the prescription is reviewed with the local dealer, at which point the farmer can choose whether to purchase the prescription, which is priced on a per acre basis ($5/acre in 2015). The farmer can then download the prescribed planting instructions for the hybrid of choice to an iPad app which will then guide the variable-rate planting equipment to plant accordingly.

Although the farmer does not have to pay until after a preview of the prescriptions is available, the farmer's data are already passed to Monsanto. The FieldScripts® program is only available through certified dealers who have an incentive to maintain their certified relationship with Monsanto and who benefit from seed and other input sales. Futhermore, only Monsanto's DEKALB® seed hybrids are available using the FieldScripts® program.[2] When the farmer accepts the prescription, she agrees to purchase prescribed seed variety at the same time, before the planting program is downloaded to the farmer's iPad. The rationale for the product bundling is that Monsanto's prescription algorithm is based on its agronomic knowledge base of its own varieties. The dealer and Monsanto's field agents help monitor performance through the season and advise on field management needs. At the end of the season, the farmer submits yield data to help improve future prescriptions for the field, which Monsanto can incorporate to update its basic algorithm as well.

Pioneer's Field360™ program and WinField's R7 program provide an even greater array of data-driven decision support tools. In addition to hybrid selection and variable-rate seed planting recommendations for specific fields, each offers additional in-season analysis for farmers using the company's seed products through web-based subscription service platforms to support farmers' crop management practices. Pioneer offers field-level weather updates so the farmer can track precipitation without having to visit remote field locations as frequently, as well as crop growth estimators based on climatic, genetic and agronomic characteristics to help the farmer identify potential deficiencies in growth that may be due to nutrition, moisture, or infestations. Pioneer's system allows the farmer not only to house all of their data in one easily-accessible location in the cloud, but to share notes (including GPS-tagged field notes) and progress data with employees, agronomy consultants, and other service providers. WinField's R7 service pairs farmers' planting and field data with satellite sensing data to help identify moisture and nutrition imbalances that allow the farmer to more efficiently apply the appropriate inputs using variable-rate systems. Like Monsanto, both operate through a network of certified dealers.

The dealer network plays a substantial role in the value chain. In their survey of commercial farm operations in the US, Alexander et al. (2009) found that 57% of respondents claimed that purchasing inputs has become a more time-consuming activity in the farm businesses. This is particularly true for the large farm operators (67%). At the same time, farmers generally trust their local dealer more than manufacturers' sales personnel and consultants as a source of information. While 50% of crop farmers claimed to be loyal to a brand of seed, 39% negatively valued information from manufacturers' salespeople and 59% negatively valued information from manufacturers' technical personnel. Roughly 2/3 of respondents claimed to value their relationship with the local dealer more than the company the dealer represents, suggesting that much of the brand loyalty reported may be driven as much by relationships with the local dealer as with the products themselves.

Dealers, then, are caught in an interesting set of incentives. Commercial farm operations want to maximize profitability per acre. For a given commodity price, that means balancing the value of the marginal product (i.e., yield) of additional inputs with the cost of those inputs. ATPs offer services that allow the farmer to increase yields and/or reduce input costs to improve profitability. The dealer is in the position of offering farmers a value-adding bundle

[2] WinField, a smaller independent seed company, announced that it would begin offering Monsanto's FieldScripts® program as a certified dealer, with plans to include options for WinField's seeds in future growing seasons in addition to Monsanto's DEKALB seed varieties.

of services that may reduce demand for the farm inputs the dealer sells. In other words, services that reduce input use by the farm cost the dealer revenue in lost sales. On the other hand, by supporting the ATP's data service programs, the dealer may benefit from a portion of program revenues and from the potential to increase margins on recommended products to compensate for reduced sales volume.

A variety of other ATPs offer less comprehensive services based on precision agriculture technologies of various types. Companies like Agrible, Conservis, and Granular provide farm management decision support and data tracking tools. Companies like AgLeader data management services to support precision planting and variable-rate chemical and moisture applications, but do not offer seed recommendations. John Deere, Case-New Holland, and AGCO are the largest of several manufacturers producing variable-rate equipment with sensors and monitors that collect data on equipment hours as well as crop input and output volumes, particularly crop yield monitors. While equipment manufacturers can use the data to provide better maintenance recommendations and performance diagnostics, they also offer data analytic software to help farmers make better use of the data that are collected.[3]

Recently, industry players have begun forming alliances to share data service resources. As noted above, WinField has contracted with Monsanto to begin offering FieldScripts®, initially limited to sales of Monsanto's DEKALB seed varieties. Monsanto, in turn, has agreed to develop FieldScripts algorithms for WinField's seed genetics, allowing WinField eventually to offer prescriptions based on their own seed varieties. Pioneer and Monsanto have each announced plans to partner with John Deere to take advantage of Deere's access to large amounts of yield monitor, and potentially other sensor, data. Such uses of farm data raise the question of who owns the data and the value created by those data's use, who has access to the data, and to what ends might it be used.

Big Data Concerns in Agriculture

In some respects, farmers' concerns about use of their farm data are no different than general consumer concerns about the security and privacy of data in the cloud: Can other people (and whom) see my data? And how are my data being used? While farm production data may not represent the identity theft risk of some consumer data, production data may be used by ag service providers not only to benefit producers by providing managerial decision support, but also to price discriminate by the nature of their recommendations. There are also concerns that data sharing rules may disadvantage farmers or at least have the appearance of creating a competitive disadvantage to local farmers.

Although farmers' concerns may not be very different than some general consumer concerns, the nature of farming and the culture of agriculture create some significant legal and policy questions. Most of the data concerns of farmers revolve around commercial use of business and business process data, not what would commonly be considered personal data privacy issues in other business sectors. Federal Trade Commission actions on data security and privacy focus are primarily based on fraud and the disclosure of "sensitive personal information." This typically includes things like names, social security numbers, bank accounts, and other financial information. Furthermore, in *Multi Ag Media LLC v Department*

[3] Given the dynamic market environment, there are undoubtedly changes in the competitive landscape in terms of specific firms and product offerings. Those selected here are simply representative and not intended to be an exhaustive listing.

of Agriculture, the D.C. Circuit Court (515 F. 3d 1224, D.C. Cir. 2008) ruled that USDA could not withhold from a Freedom of Information Act request information on "irrigation practices, farm acreage, and the number and width of rows of tobacco and cotton" or GIS database information "on farm, tract, and boundary identification, calculated acreage, and characteristics of the land such as whether it is erodible, barren, or has water or perennial snow cover." Although the Court did find that there was a privacy interest in the data, it wrote in its option that "we are not persuaded that the privacy interest that may exist is particularly strong" and ruled that the public interest in releasing the data outweighed any privacy concern. Consequently, federal policies protecting personal information likely would not apply to the kinds of data collected by ATPs.

However, the legacy and culture of "family farm" operations create an interconnected sense of identity between farmers and their farm businesses. Traditionally, the farm business and farm household were viewed as one-and-the-same economic unit, as production and consumption decisions were integrally intertwined (Heady et al. 1953; Brewster 1979). While Mishra et al. (2002) argue the economic portfolios of farm households no longer reflect such reliance and interdependence, the intertwined identity of household and farm business remains. Pritchard et al. (2007, 2) argue:

> "[T]he 'typical' family farm is now far removed from the ideal-typical family farm origins; yet neither is it akin to a corporate farm. Rather, it represents a distinct social and economic formation in its own right. … Farming may well be becoming more corporatized, but it also retains distinctive social properties (based mainly around family ownership) that separate it conceptually from other segments of the economy."

So while farm data conceptually may be akin to—and perhaps subject to the same legal treatment as—commercial data, in the minds of farm producers these data are often viewed with a personal sense of privacy concerns. That said, it is not clear that agricultural producers are at any more risk of data misappropriation or value risk than the data service providers themselves.

Farm-Level Data Issues

A primary concern of farmers at the farm level is that the data collected and the recommendations generated by ATPs may facilitate what economists refer to as first-degree price discrimination by input suppliers, with customized pricing based on farm attributes.[4] This could be from ATPs who sell inputs themselves, or input suppliers that may partner with or purchase data from other ATPs. For instance, Pioneer's Field360™ is limited to farmers using Pioneer seed. FieldScripts® is limited, at present, to Monsanto's DEKALB seed varieties. The bundling of data services and specific input brands gives rise to concerns that the ATP may increase input costs to capture the anticipated gains from more efficient production resulting from the prescribed practices. As discussed above, dealers have incentive to increase margins on seeds that are prescribed, or eligible to be prescribed, particularly if more efficient planting practices reduce overall seed demand. However,

[4] Block price (third-degree) price discrimination is nothing new in agriculture, as large farmers may receive bulk discounts to lower the average cost per unit of input. First-degree price discrimination, however, targets the price discrimination to individual buyer characteristics, not simply quantities. This allows the seller to extract even more of the gains from trade from each prospective buyer.

because these services are only now rolling out to large regional markets, there are no data available to substantiate or dispel price discrimination concerns.

Farmers could induce competition to mitigate the potential discrimination by contracting with more than one ATP for different fields, particularly large farm operations. However, as Alexander et al. (2009) found, farmers tend to be loyal to specific dealers and input brands. Moreover, working with multiple ATPs for inputs—particularly for seed—would increase learning costs of using the different systems in different fields. Finally, the pricing of services may also reduce the incentive to use multiple vendors. While FieldScripts® charges by the acre, Field360™ is a flat annual subscription which can be averaged out across larger numbers of acres. Given that complexity of input decision making is one of the motivations for using these data services, using multiple platforms would seem counter-productive.

Given the general skepticism farmers seem to hold against input manufacturers (Alexander et al. 2009), it is not surprising that farmers are concerned about manufacturers using the farm's data to create additional value beyond the farm without compensating the farmers for the data and to potentially gain competitive advantage over other potential industry entrants. ATPs use client farms' performance (yield) data to improve their algorithms, much like Google, Amazon or Facebook improve the value of their algorithms based on individual users' searches and consumption decisions. Given the annual nature of growing seasons and the rate of change in production technologies, data aggregation across a large number of farms is critical to developing improved crop forecasting models and production recommendations. Individual farm-level longitudinal data would not generate sufficiently large data samples to estimate useful models. This gives rise to concerns about the ability of large incumbent ATPs to hold a competitive advantage over any potential entrants that do not have access to the breadth of data (i.e., data from a large number of farms).

As in the case of Google, Amazon or Facebook, it is difficult to imagine that any one farmer's data has material value at the margin in the development of data analytic algorithms. However, unlike the large internet companies that have millions of users providing data, Monsanto may have only tens of thousands. Moreover, increased consolidation in the operation of farmland means a smaller number of farmers operate a large percentage of the acres being farmed. Thus, the marginal value of a particular large farming operation to a company like Monsanto, while small, may represent a significantly larger share of the company's data than any one user of Google, Amazon or Facebook. This is especially true given the geo-climatic-specific nature of farms, since soil characteristics and climate both vary regionally and are key determinants of crop production and farm input choice. The attributes of large individual farming operations—or a group of farms in a particular region—may add variation to the estimation sample that allows for more precise estimates of key parameters. Thus, while it is unlikely a relatively small group of consumers could collectively offer Google or Facebook a block of data with a meaningful marginal value, the same may not be true in the case of farm-level production data. Thus, the question of data ownership would seem to have competitive consequences in the ability of large growers or groups of growers to market their data to potential entrants to the ag data services industry.

Data Issues Beyond the Farm

While consumers' concerns focus on personal privacy and, potentially, price discrimination by online retailers for things the consumer wants to buy and how those retailers use their data for developing their products, farmers' concerns extend further. In addition to personal privacy issues and concerns about farm input providers potentially engaging in price discrimination for seeds and chemicals, farmers are also concerned about data aggregators using the data (or making the data available for others to use) to gain an unfair advantage in commodity and real estate markets, which have significant implications for the value of farm operations. While not strictly of a personal nature, these reflect concerns about the privacy of data that has been aggregated being used for alternate purposes.

For instance, data from yield monitors mounted on harvesting equipment may now be transmitted into the cloud in near-real time, giving the data aggregator real-time crop yield data for specific, identifiable tracts of farmland—across large numbers of farms at any given time. Companies like John Deere ostensibly use that data to diagnose their harvesting equipment's performance and recommend routine maintenance based on machine use. Precision ag service providers such as AgLeader® not only capture harvest data, but planting, application and irrigation data as well, capturing several of the key, controllable input factors of production to provide management decision support to producers. However, access to such data could be used to speculate in commodities markets with information that is not otherwise knowable to market participants, giving rise to concerns about market manipulation.

Similarly, aggregated yield data may be used (or sold to others to use) to identify the most productive tracts of farmland on a national scale, further facilitating investments by REITs into the agriculture real estate market. Farmland has been described on Wall Street as "gold with a coupon," reflecting the fact that farmland prices have historically appreciated on a consistent basis in addition to generating annual revenues (Moyer 2014). Markets and competition for farmland have traditionally been more regional between active farmers and landowners, many of whom are retired farmers. Increased competition by non-farming investors puts pressure on land prices to increase. While farming landowners' balance sheets benefit from increasing land values, high land prices are frequently cited as the primary barrier to entry for new and young farmers who do not have the resources to acquire land on a scale to allow economically efficient crop production.[5] Moreover, farmers' cash flows are negatively affected by increased land prices and cash rental rates. According to the 2012 Census of Agriculture, approximately 40% of farmland nationally is operated under lease or rental agreements, with much higher percentages in the Midwest grain belts where big data issues are perhaps of greatest relevance—and concern, as evidenced in part by the farmer associations party to the data privacy principles agreement noted above.

Finally, farmers may be concerned about government collection and use of data, either for farm program analysis or for environmental enforcement. A recent paper by Antle et al. (2015) arguing for the utility of a new data infrastructure to exploit big data for agro-environmental policy analysis would seem to justify concerns about such uses. Antle et al. recognize that such a program would likely have to be voluntary to be "politically and socially acceptable" (2015, 5) in light of farmers' expressed concerns about data privacy and security concerns, but the authors also explain the potential short-comings of a voluntary (versus mandatory) system for statistical analysis purposes. Compound traditional privacy concerns with the possibility of Environmental Protection Agency (EPA) access to data on

[5] An economically efficient crop operation may require a minimum of 500 to 1,000 acres. Even at $5,000 per acre, entry into farming would require $2.5 to $5 million dollars just in land. Thus, a common quip that the only way for a young adult to get into farming is by death (inheritance) or marriage.

farmers' pesticide and herbicide application practices, and farmers' concerns about data access and use is not surprising.[6]

Thus, farmers have a wide range of concerns about the uses of their data and how those uses may end up putting farmers at a competitive disadvantage relative to the companies with whom they are sharing their data to begin with.

Turning Data Concerns on Their Head

While the concerns about big data in agriculture have primarily been cast from the farmer's perspective, it is also of concern to the ATPs whose algorithms embody the information generated by using those data and whose outputs reflect the nature of those algorithms. This creates a potential tension in the ownership and use of farm data. Again using Monsanto's FieldScripts® program, if the farmer owns the data on their actual planting rates and the field map data that Monsanto used to generate the farmer's prescriptions, one could imagine those data being used by competitors to backward-out Monsanto's prescription algorithm, or a close approximation. However, farmers who may want to work with a different ATP in subsequent years—or ones who work with multiple ATPs in a given growing season—may have need to share data with them in the course of using their services. Likewise, ATPs may have concerns about receiving data from farmers that the farmer herself does not own, giving rise to potential violations of intellectual property or licensing restrictions. As noted above, ATPs are increasingly forming alliances or partnerships with each other to access one another's data or knowledge bases. While on the surface these arrangements may be intended as value-adding for the various participants, the arrangements may also be defensive in nature to prevent disputes over data ownership and use.

Principles of Data Privacy and Security in Agriculture

Given the attributes of and issues associated with big data in agriculture discussed above, it is clear that both sides of the data equation have an interest in developing clear property rights over agricultural production data and its use. In November 2014, several agricultural producer organizations and leading ATPs announced agreement on a set of principles to govern data use and sharing.[7] The principles outline an agreed upon approach to dealing with data issues and an agreement to continuing dialogue as new technology and data issues evolve.

In terms of ownership, the principles state that farmers retain ownership of "information generated on their farming operations." This definition creates a clear delineation between data or information generated using the farmer's data and information generated on the farm itself. Thus, the principles suggest a farmer would not own data reflecting the recommendations of ATPs, such as planting guides and rates, despite those planting data being used on the farm operations.

[6] One might just as reasonably argue that application data may provide a defense against EPA claims. However, that would be equally valid without giving *a priori* access to the EPA to use in identifying potential sources of violations.
[7] The participating organizations include American Farm Bureau Federation®, American Soybean Association, Beck's Hybrids, Dow AgroSciences LLC, DuPont Pioneer, John Deere, National Association of Wheat Growers, National Corn Growers Association, National Farmers Union, Raven Industries, The Climate Corporation – a division of Monsanto, and USA Rice Federation.

The principles further state "it is the responsibility of the farmer to agree upon data use and sharing with the other stakeholders with an economic interest, such as the tenant, landowner, cooperative, owner of the precision agriculture system hardware, and/or ATP, etc." In the case of owners of precision agriculture system hardware and ATPs, data must be shared between the farmer and the service provider as a fundamental aspect of the transaction. For data sharing between the farmer and landowner, tenant, or cooperative, there is no *a priori* reason that data should be shared as a matter of principle. Landowners, farmers or tenants may have reason to withhold information about their productive efforts either in negotiating or implementing cash rent or share contracts.

Allen and Lueck (1992, 1993) argue that the choice of cash rent or share contract, and the amount of the share, are based in part on the information asymmetry between farmers and landowners. Farmers have incentive to exploit soil quality, underinvest in maintaining or improving soil quality, and misreport input use and output yields to landowners from whom they rent. Depending on the type of farm operation (e.g., irrigated or non-irrigated; more or less volatile annual yields, etc.) and the cost of measurement, the optimal form of contract (share vs. rent) and the amount of share or size of rent payments would be different. As new technologies improve the ability to objectively verify input use and yields, one would expect parties to request access to this data as part of their contract negotiations and for contract terms to reflect the reduced information asymmetry between parties. But one would expect these terms to evolve naturally as a matter of competition between potential tenants and landowners.

The principles statement further stipulates the expectation of privacy, informed consent, transparency, and the ability to choose the level and type of information that can be collected and used by ATPs. The agreed upon practices mimic the kinds of privacy agreements seen among most online service providers' end-user license agreements (EULAs): contractual consent documents, including disclosures of how data will be used and the third parties with whom data will be shared, "whether signed or digital." This means producers could be presented with a standard privacy agreement as part of the ATPs EULA and asked to "Click Agree" to continue, thereby consenting to the privacy terms without having to physically read and sign the document. The principles further state the contract should also outline the ATP's options for farmers to limit use or disclosure of their data, and should provide for famer's individual, identifiable data to be retrieved from the ATP's records and returned to the farmer in a format that is portable or compatible with other data systems, allowing the farmer to share it with an alternate ATP.

Farmers are also given *de jure* residual control rights over the use of their data beyond the original terms of use with the ATP. The principles state that "[a]n ATP will not sell and/or disclose non-aggregated farm data to a third party without first securing a legally binding commitment to be bound by the same terms and conditions as the ATP has with the farmer. Farmers must be notified if such a sale is going to take place and have the option to opt out or have their data removed prior to that sale." On the surface, this allocation of rights would seem to protect farmers' privacy interest and would make the sale of data to third parties extremely costly for ATPs who have to obtain consent from every farmer customer. However, given the data from any single farm is unlikely to be of much value to a third party, the *de facto* rule is that ATPs can sell or disclose aggregated data to any third party that is willing to abide by terms similar to the ATPs original consent disclosure agreement. Notably, this provision does not distinguish between aggregated and farm-identifiable data, as with the farmer's retrieval policy. While the latter specifically refers to data "that has been made anonymous or aggregated and is no longer specifically identifiable," the residual control rights apply only to "non-aggregated farm data." Consequently, the terms offer little meaningful protection for farmers' privacy. However, this is probably in farmers' best interests as a whole.

Finally, the data privacy principles specifically prohibit the use of data "for unlawful or anticompetitive activities, such as a prohibition on the use of farm data by the ATP to speculate in commodity markets." Obviously, a prohibition against unlawful use is superfluous, except insofar as to frame the restriction on behavior that may be deemed anticompetitive but not otherwise be unlawful. Speculative trading in agricultural commodities is subject to regulatory limits imposed by the Commodity Futures Trading Commission's (CFTC) Regulation 150.2. The CFTC's regulations are based on volume limits, regardless the source of a speculator's information, since it is the size of positions that moves market prices. Consequently, it is unclear whether the concerns around speculative trading using farm data are well-founded and whether prohibitions against otherwise-legal speculative trading would in fact benefit farmers. I'll return to this below.

Competitive Implications of Ag Data Ownership and Privacy

The Principles statement is an attempt by industry to implement a privately ordered set of norms in the absence of (or to preempt) specific formal regulations governing ag data privacy and security. The questions, however, are whether any such norms are necessary and whether the proposed restrictions—much less any federally-imposed restrictions—may themselves result in anticompetitive effects and a reduction in competition and innovation in this growing industry for ag data services. In this section, two areas of competitive concern are addressed: restrictions on data transfer, and restrictions on use of data beyond the farm.

Data Ownership and Restrictions on Data Sharing

The proposed data ownership and sharing rules limit the ability of farmers to share all the data they have available from their farming operations with potential service providers, potentially limiting the quality of services that can be offered from competing ATPs. While farmers own data generated on their farming operations, such as soil maps, weed maps, and harvest data, application data for seeds or other inputs may not belong to the farmer if those data were prescribed and provided by ag data service companies. In cases like Monsanto's FieldScripts® program, in which application rate and guidance data are provided on a tablet device that is simply plugged into the farmer's (or farm equipment owner's) machinery, the farmer may never have access to the precise application data itself unless that data can be downloaded to the farmer's computer. At best, the farmer would know only the total amount of inputs applied over a given number of acres (average application rates).

Since the prescription data does not belong to the farmer, the farmer cannot transfer the data to a third party ATP outside the relationship with the generating ATP. Moreover, contractual restrictions may further limit the ability of the farmer to use the data for herself. As an example, Monsanto's service agreement specifically states that "FieldScripts® and the related algorithms and documentation are the intellectual property and proprietary information of Monsanto. Grower may not transfer FieldScripts® and its related information to any third party for reverse engineering FieldScripts®. This provision shall survive termination of this Agreement." In addition to limiting the sharing of data with third parties, the service agreement also prohibits farmers from using the FieldScripts® data on fields not contracted for the program, effectively limiting the farmer's use of data that may traditionally have been owned and used by the farmer for her own analytics.[8]

While the principles statement stipulates that farmers own their own data, the guidelines limiting ATP's ability to share that data leave open a wide range of uses provided they are consistent with the ATP's original data privacy and collection notice. In theory, the ATP could license use of the data to other ag data service providers to develop technologies and algorithms that would enhance or expand the ATP's product and service portfolio. While this would create an opportunity for new entrants to gain access to data, it clearly places large, incumbent ATPs in a roll of gatekeeper for new product development that relies on historical farm data.

The competitive advantage of the ATP incumbent when it comes to potential new market entry is twofold. First, the large ATPs have access to a broad cross-section of farm data aggregated across their many farm customers. As argued above, this is likely more important than a long history of farm-level data. Sonka (2014) argues that the idiosyncratic nature of agricultural production resulting from the biological production process limits the salience of older data for forecasting future yields. Add to that, changes in agricultural technologies and the limitation of annual observations, and it quickly becomes evident that cross-sectional variation over a limited number of seasons provide better data for identifying relationships between important input factors and practices. Second, given the limitations on the kinds of data farmers can share or may even have available from their operations, potential entrants may not be able to acquire as much data directly from farmers as would be available to (and from) the ATP, since the ATP is able to use both the farmer's data and its own in refining its algorithms and developing new products and services.

Given the relative novelty of big data and its competitive effects, the appropriate role of antitrust law to mitigate market foreclosure concerns is unsettled. Geradin and Kuschewsky (2013), Newman (2014), and Stucke and Grunes (2015) all argue that competition law can (and should) play an important role in mitigating the entry barrier effects of large first-movers in big data sectors. Manne and Rinehart (2013) argue that argument about market foreclosure failure to understand the economics data-driven businesses. However, these papers all focus on consumer personal data in the context of social media platforms that are offered to consumers at a price of zero, which raises the question of what anticompetitive harm would even mean. That is not the case in the ag data services industry, which thus far has not received any attention in the antitrust arena. That said, the FTC follows a rule of reason approach in enforcing such anticompetitive concerns under the Robinson-Patman Act, meaning the economic efficiencies of data ownership for innovation and product development among ATPs would be weighed against any potential market foreclosure and evidence of competitive harm to farmers.

If farmers are concerned about potential ability of ATPs to foreclosure market access to new entrants, they have the ability to reduce transaction costs for potential entrants while also increasing the value of their individual data by forming agricultural data marketing cooperatives. Unlike most any other business sector, agricultural producers have exemption from antitrust laws in jointly marketing their products, including data resources. Because no one farmer's data adds significant value at the margin, autonomous data markets would result

[8] I did not find terms of use for Pioneer and WinField, but it is reasonable to believe they would include similar use restrictions.

in near-zero prices for an individual farm's data. By aggregating data from a large number of farmers, the value of the data resource would be larger and generate larger average returns than could be achieved individually.

Grower Information Services Cooperative (GISC) was originally formed to aggregate farm data and sell it to crop insurance companies to improve their risk management and pricing strategies. Since Monsanto's acquisition of The Climate Corporation in 2013, GISC has broadened its scope and marketing to potential cooperative members as an alternative mechanism to share in the value of their data for developing ag data services. Farmer cooperatives continue to control a significant share of agricultural input retail and contract services for their customer-owners. These cooperatives already have access to large amounts of data from their members, but have thus far not shown a recognition or ability to take advantage of that data access to capture value for their members by marketing their collective data resources. A limiting factor may be uncertainty over what information can legally be collected, packaged, and resold given their members' existing arrangements with ad data service ATPs. Another may be the fact that these local cooperatives are often certified dealers for the ATPs themselves.

Restrictions on Use of Data Beyond the Farm

As noted above, concerns about the use of data beyond the farm extend beyond the farm operations themselves to the ability to use that information in commodity and real estate markets. Concerns about selling data to institutional investors and REIT managers could be covered by limitations on transferring data for uses beyond the scope outlined in the ATP's original data collection agreement. However, use of data for "unlawful or anticompetitive activities" is more specifically targeted to manipulation of commodity markets. As noted earlier, the CFTC restricts any party from what it defines as excessive speculation in the market (17 CFR 150.2), including specific position limits on corn, wheat and soybeans (70 FR 24706, May 11, 2005). The limits are based on average contract volumes for each commodity's futures contracts so as to limit the ability of any one speculator from cornering or manipulating the market—regardless the source of their information.

There are two principle periods in which ATPs might have a significant information advantage based on their access to farm-level data: planting and harvest. At harvest, real-time yield data could give data aggregators earlier insights on expected harvest yields than are otherwise available. Currently, reliable yield data are not readily available in the market until sometime after harvest. Elevators may report total deliveries, but do not know the actual number of acres harvested nor how much of the harvest may have been stored on-farm. A small number of private companies also collect data from elevators and conduct field surveys to provide private crop reports. The USDA also produces harvest forecasts based on various survey instruments and data collection efforts. At each stage, there is a delay in information being made available to market participants. Real-time harvest yield data could give the aggregator a lead of anywhere from days to weeks. However, near-month contracts have the strictest regulatory limits on speculative positions; only 600 contracts in any of the major cash crops. Even if a data aggregator attempted to use its information advantage, its ability to engage in anti-competitive trading at harvest would be severely limited by existing trade limitations.

Planting data may provide a greater opportunity to engage in speculative behavior since position limits are significantly larger for longer-term positions.[9] As ATPs gather data on the number of acres planted and, for companies that also perform crop forecasting like Monsanto, projections of yields for those acres based on agronomic and climatic conditions and forecasts, they would have an advantage in taking on speculative positions. This would continue through the growing season as in-season data are collected and compared against growth prediction models, like Pioneer's Field360™ model.

However, such speculative trading provides information to the market—and for this reason; the information is arguably superior to what is currently available. This would make futures prices more accurate predictors of future prices, which would allow farmers to more accurately hedge and price the value of their crops. With increasing global demand for agricultural commodities, increasing ties between agricultural commodity prices and the energy complex through biofuels, and structural changes like moves to electronic trading, commodity prices have become more volatile. Some have raised concerns about the quality of price discovery as a result of these changes (Irwin and Sanders 2011). Allowing data aggregators to trade on their data would actual improve market conditions for all interested parties by bringing more accurate and scientifically grounded information into the market. On this point, the principles for data privacy and security negotiated by the industry are likely to work against farmers' interests.

Conclusion

As noted in the introduction, Whitacre et al. (2014, 3), argue that "[b]efore 'big data' will be widely accepted by farmers and others across the agricultural industry, its collection, processing, on-demand analytics, and decision making must become passive to the user." Agricultural technology and data service providers are quickly moving into this space to provide decision making tools for farmers that require very little analytic skill or active participation by the farmer. Data can be directly uploaded from the farm equipment to the provider's servers for analysis. The provider can return a tablet that simply plugs into the farmer's equipment and applies the results of the ATP's analysis and prescriptions. We are just on the front edge of a data-driven transformation of the agricultural sector.

But with such innovation comes other concerns about a data-driven industry: who owns the data, how can they use it, and what does that mean for competition and industry dynamics? If Whitacre et al. were concerned that farmers needed *'big data'* to be served up for them, they overlooked the privacy and competition concerns of farmers about the use of their data. Industry players have attempted to outline a guiding set of principles concerning data ownership. Proposed answers to questions about ownership and use raise concerns about the implications for new market entrants and commodity market performance. An analysis of the proposed industry standards suggests that farmer groups are likely worried about the wrong things, specifically with regard to data use for speculation, rather than the dynamic consequences of data usage guidelines for future competitive market outcomes.

[9] For instance, the aggregate position limits for corn, wheat and soybeans are 33,000 contracts, respectively, compared to spot month limits of 600 contracts.

Acknowledgements

The author wishes to acknowledge support for this project from the George Mason University Law & Economics Center's 2015 Privacy Fellows program.

References

Alexander, C., M. Boehlje, S. Downey, A. Gray, M. Gunderson, and M. Roucan-Kane. 2009. Serving Producers in Volatile Times: Report from the 2008 Large Commercial Producer Survey. Working Paper 09-09, Purdue University. http://purl.umn.edu/56116.

Allen, D. W. and D. Lueck. 1992. Contract Choice in Modern Agriculture: Cash Rent versus Cropshare. *Journal of Law and Economics* 35(2): 397–426.

________. 1993. Transaction Costs and the Design of Cropshare Contracts. *The RAND Journal of Economics* 24(1):78–100.

________. 1998. The Nature of the Farm. *Journal of Law and Economics* 41(2): 343–386.

Ali, Sarah. 2014. Directory of Rural—Farmer, Rancher, and Fishery Cooperatives. *Rural Development Service Report 22*, U.S. Department of Agriculture.

Antle, J., S. Capalbo, and L. Houston. 2015. Using Big Data to Evaluate Agro-environmental Policies. *Choices* 30(3). http://www.choicesmagazine.org/UserFiles/ file/cmsarticl _ 465.pdf.

Brewster, D.E. 1979. The Family Farm: A Changing Concept. Structure Issues of American Agriculture. U.S. Department of Agriculture, Economics, Statistics, and Cooperative Service, AER-438. http://naldc.nal.usda.gov/naldc/download.xhtml?id=CAT84802668&content=PDF.

Geradin, D., and M. Kuschewsky. 2013. Competition Law and Personal Data: Preliminary Thoughts on a Complex Issue. SSRN: http://ssrn.com/abstract=2216088 or http://dx.doi.org/10.2139/ssrn.2216088.

Griffin, T.W., J. Lowenberg-DeBoer, D.M. Lambert, J. Peone, T. Payne, and S.G. Daberkow. 2004. Adoption, Profitability, and Making Better Use of Precision Farming Data. Department of Agricultural Economics, Staff Paper #04-06, Purdue University, West Lafayette, IN.

Heady, E.O., W.B. Back, and G.A. Peterson. 1953. Interdependence Between the Farm Business and the Farm Household with Implications on Economic Efficiency. Research Bulletin 398, Department of Economics and Sociology, Iowa State College, Ames, IA.

Irwin, S.H., and D.R. Sanders. 2011. Index Funds, Financialization, and Commodity Futures Markets. *Applied Economic Perspectives and Policy* 33(1): 1–31. doi: 10.1093/aepp/ppq032.

Manne, G.A., and W. Rinehart. 2013. The Market Realities that Undermined the FTC's Antitrust Case against Google. *Harvard Journal of Law & Technology* 26(2):(Spring).

Mishra, A.K., El-Osta, H.S., Morehart, M.J., Johnson, J.D., Hopkins, J.W. 2002. Income, wealth, and the economic well-being of farm households. Farm sector performance and well-being branch, Resource Economics Division, Economic Research Service, US Department of Agriculture. Agricultural Economic Report No. 812. U.S. Department of Agriculture, Washington DC.

Moyer, L. 2014. Time to Buy the Farm? *Wall Street Journal Online* 7-September. http://www.wsj.com/articles/reits-give-small-investors-access-to-farmland-and-forest-tracts-1410120112. Accessed March 24, 2015.

Newman, N. 2014. Search, Antitrust and the Economics of the Control of User Data. *Yale Journal on Regulation* 30 (3). http://papers.ssrn.com/sol3/papers.cfm? abstract_id= 2309547.

Pritchard, B, D. Burch, and G. Lawrence. 2007. Neither 'family' nor 'corporate' farming: Australian tomato growers as farm family entrepreneurs. *Journal of Rural Studies,* 23:75–87.

Schimmelpfenning, D. and R. Ebel. 2011. On the Doorstep of the Information Age: Recent Adoption of Precision Agriculture. U.S. Department of Agriculture, Economic Research Service, Washington, D.C.

Stucke, Maurice E. and Allen P. Grunes. 2015. Debunking the Myths Over Big Data and Antitrust. *CPI Antitrust Chronicle* University of Tennessee Legal Studies Research Paper No. 276. SSRN: http://ssrn.com/abstract=2612562.

Sonka, S. 2014. Big Data and the Ag Sector: More than Lots of Numbers. *International Food and Agribusiness Management Review* 17(1):1-20.

Whitacre, B.E., T.B. Mark, and T.W. Griffin. 2014. How Connected are Our Farms? *Choices* 29(3). www.choicesmagazine.org/UserFiles/file/cmsarticle_392.pdf.

Zhang, C. and J.M Kovacs. 2012. The application of small unmanned aerial systems for precision agriculture: A review. *Precision Agriculture*. 13(6): 693-712. doi: 10.1007/ s11119-012-9274-5.

International Food and Agribusiness Management Review
Special Issue - Volume 19 Issue A, 2016

Targeting Drought-Tolerant Maize Varieties in Southern Africa: A Geospatial Crop Modeling Approach Using Big Data

Kindie Tesfaye[ⓘa], Kai Sonder[b], Jill Cairns[c], Cosmos Magorokosho[d], Amsal Tarekegn[e], Girma T. Kassie [f], Fite Getaneh [g], Tahirou Abdoulaye [h], Tsedeke Abate [i], and Olaf Erenstein [j]

[a g] *International Maize and Wheat Improvement Center (CIMMYT), ILRI Campus, CMC Road , P.O. Box 5689, Addis Ababa, Ethiopia*

[b, j] *CIMMYT, El Batan, Mexico*

[c, d, e] *CIMMYT, Harare, Zimbabwe*

[f] *International Center for Agricultural Research in the Dry Areas (ICARDA), Addis Ababa, Ethiopia;*

[h]*International Institute of Tropical Agriculture (IITA),*

[i] *CIMMYT, Nairobi, Kenya*

Abstract

Maize is a major staple food crop in southern Africa and stress tolerant improved varieties have the potential to increase productivity, enhance livelihoods and reduce food insecurity. This study uses *big data* in refining the geospatial targeting of new drought-tolerant (DT) maize varieties in Malawi, Mozambique, Zambia, and Zimbabwe. Results indicate that more than 1.0 million hectares (Mha) of maize in the study countries is exposed to a seasonal drought frequency exceeding 20% while an additional 1.6 Mha experience a drought occurrence of 10–20%. Spatial modeling indicates that new DT varieties could give a yield advantage of 5–40% over the commercial check variety across drought environments while crop management and input costs are kept equal. Results indicate a huge potential for DT maize seed production and marketing in the study countries. The study demonstrates how big data and analytical tools enhance the targeting and uptake of new agricultural technologies for boosting rural livelihoods, agribusiness development and food security in developing countries.

Keywords: big data, drought tolerance; geospatial analysis; maize; spatial crop modeling, targeting

ⓘCorresponding author: Tel: +251.116.462324
Email: K. Tesfaye: K.tesfayefantaye@cgiar.org

Introduction

Rain-fed agriculture produces much of the food consumed globally and provides for the livelihoods of rural communities across the developing world. It accounts for more than 95% of farmed land in sub-Saharan Africa (SSA) where the rural populace of predominantly resource-limited families still face poverty, hunger, food insecurity and malnutrition (Wani et al. 2009). Maize is the most important staple food crop in SSA where it is almost entirely grown under rain-fed systems which are dependent on increasingly erratic rainfall. In southern Africa, maize accounts for 77% of the cereal area and 84% of the production, and over 30% of the total calories and protein consumed (FAOSTAT 2015).

However, current maize production in SSA is not sufficient to meet the growing demand in most countries and yields remain among the lowest in the world (Ray et al. 2012) because of an array of biophysical and socioeconomic constraints (Shiferaw et al. 2011). Drought is one of the major constraints under rain-fed systems with an estimated 40% of SSA's maize area facing occasional drought stress causing a yield loss of 10–25%. Around 25% of the maize crop suffers frequent drought resulting in a loss of up to half the harvest (CIMMYT 2013a). In southern Africa, maize yields are typically low due largely to drought and low-N stress (Weber et al. 2012).

Enhancing the productivity of rain-fed agriculture is an important avenue in reducing poverty and food insecurity in rain-fed systems (Rockström and Barron 2007; Wani et al. 2009). For example, adoption of improved maize varieties increases productivity and reduces chronic and transitory food insecurity under rain-fed systems (Kassie et al. 2014). Thus, increasing the use of improved technologies has the potential to enhance the welfare and food security of poor households (Bezu et al. 2014; Kassie et al. 2014). Improved maize technologies have been developed, disseminated and made positive contributions to the livelihood of smallholder farmers in some African countries (e.g., Abate et al. 2015). However, increasing adoption among smallholder farmers in Africa remains a challenge, including for DT maize varieties (Fisher et al. 2015). One of the challenges for wider adoption is the lack of data and tools for targeting new technologies at scale. Targeting is defined here as a process of identifying where a particular technology is the most likely to be successful–i.e. pinpointing the technology geo-spatially to the most likely niches of success. Targeting does not ensure the technology will be adopted there, but it does provide an indication of a potential fit between technology supply and demand in a geo-spatial context; and it is closely associated with recommendation domains (Notenbaert et al. 2013; Tesfaye et al. 2015c). In the context of targeting, data generated from a few research stations and/or on-farm demonstration plots are often not representative enough to address spatial and socioeconomic heterogeneities across scales.

Lately, climate, soil, elevation, and vegetation data sets are widely available at different spatial scales supporting analyses that were much more difficult in the recent past (Hyman et al. 2013). Big data and predictive analytics can make a difference in the agricultural industry (Sabarina and Priya 2015). Crop improvement and adoption research and development efforts have already benefitted from advances in big data, computing technology, and crop modeling for targeting genotypes to diverse environments (Löffler et al. 2005; Hyman et al. 2013). Targeting of crop varieties using a combination of big data and analysis tools has generated interest from public and private seed companies who wish to verify the area of adaptation and the agronomic value of new varieties for planning proper seed marketing and advisory schemes (Annicchiarico 2002). Therefore, the objective of this study is to assess the potential of targeting new DT maize varieties in southern Africa based on adaptation and productivity gains of new DT maize varieties, and present policy implications for seed production planning, marketing, and/or adoption. The study employs geospatial analysis and crop modeling tools that handle high resolution gridded climate, soil and crop data. The study purely focuses on the prospective technology change of using seed of a new DT maize hybrid instead of the prevailing non-DT commercial hybrid seed in areas that already produce maize—keeping other inputs constant. The study, therefore, does not include other productivity enhancing or risk-reducing interventions (be it crop rotation, crop management, and/or input considerations) nor does it assess the general suitability for maize in the study regions or its comparative advantage. The study contributes to a growing field of targeting research to inform agricultural development opportunities–typically linked to specific technologies and agro-ecological characteristics (Homann-Kee Tui et al. 2013; Hyman et al. 2013; Notenbaert et al. 2013; Tesfaye et al. 2015c) and/or socio-economic characteristics (Erenstein et al. 2010; Lang et al. 2013).

Methodology

Study Region

The study was conducted in four major maize-growing countries (Malawi, Mozambique, Zambia and Zimbabwe) in southern Africa. In these countries, maize stands out as the primary crop in terms of area, absolute yield levels, and staple source of food (both calorie and protein) for millions of households (Kassie et al. 2013). Maize production in the region is constrained by several biophysical and socioeconomic factors. Amongst the biophysical factors, drought stands out as the major challenge across the region (Kassie et al. 2012; Weber et al. 2012). The study area is comprised of six Maize Mega–Environments (MME): dry lowland, wet lowland, dry mid-altitude, wet lower mid-altitude, wet upper mid-altitude and highland. MMEs are areas with broadly similar environmental characteristics for maize production delineated using environmental factors (maximum temperature, rainfall, and soil pH) as explanatory factors in capturing genotype by environment interactions (Hodson et al. 2002).

Dataset for Geospatial Drought-Frequency Analysis

The frequency of drought occurrence in the maize-growing environments of the study countries during the main cropping season (October–April) was analyzed using a long-term (1960–1998) gridded (0.5 x 0.5 degrees) standardized precipitation index (SPI) calculated using the climate database of the University of East Angelia (UEA) (Mitchell and Jones 2005). The SPI values were downloaded from the online database of the International Research Institute for Climate and Society (IRI 2015). The SPI simply refers to the number of standard deviations that an observed cumulative precipitation deviates from the climatological average (Mckee et al. 1993). The focus of our analysis was on seasonal drought and hence the six–month SPI values used for the study were for the period from November to April, which is the main rainy season in southern Africa.

Geospatial Drought-Frequency Analysis

The SPI values can be classified into three wet (SPI $\geq$ 1), three dry (SPI $\leq$ -1) and one normal (1>SPI>-1) classes (Sienz et al. 2012). For simplicity and ease of presentation, the study focused on the frequency of drought occurrence rather than comparing drought severity.

Therefore, pixels with values of ≤ -1 were classified as drought years while those with values of > -1 were classified as non-drought years. The frequency analysis was done using the 'equal to frequency' tool in ArcGIS 10.2 software (http://www.esri.com). The tool evaluates the number of times a value in a set of rasters is equal to a reference value raster (drought or non-drought in this case) on a cell-by-cell basis. Therefore, for each cell location in the input reference value raster, the number of occurrences where a raster in the input list has an equal value is counted. This was then converted to percentage frequency that explains the probability of occurrence of a drought or non-drought year for each pixel. A geospatial analysis was used to map and calculate the areas under different drought frequencies (1–10%, 10–20%, 20–30%, and >30%) across the six MMEs.

Spatial Crop Modeling

A spatial crop-modeling framework that integrates climate, soil, crop and crop management data was used to assess the performance of new DT maize varieties across environments in southern Africa.

Model Description

The Cropping System Model (CSM) used for simulating maize yields was Crop Estimation through Resource and Environment Synthesis, CERES–maize (Jones and Kiniry 1986), which is embedded in the Decision Support System for Agrotechnology Transfer (DSSAT), Version 4.5 (Hoogenboom et al. 2010). CERES–maize is a process-based, management-oriented model that utilizes water, carbon, nitrogen and energy balance principles to simulate the growth and development of maize plants within an agricultural system. The model runs with a daily time step and simulates crop growth, development and yield of specific cultivars based on the effects of weather, soil characteristics and crop management practices (Jones et al. 2003).

Genetic and Environmental Data for Model Calibration and Evaluation

Five new DT maize hybrids (CZH0946, CZH0811, CZH0616, CZH0835, and CZH0837) which represent four different maturity groups (extra-early, early, medium and late maturing) and one commercial check hybrid (SC513) that is widely grown in the region were selected for the study. The new hybrids are developed for southern and eastern Africa through a rigorous breeding specifically for yield potential and yield stability in drought-prone environments (Cairns et al. 2013). The CERES-Maize model was calibrated and evaluated using long-term (2005–2011) field data collected from a network of DT maize experiments in southern Africa, particularly from Zimbabwe. Data on crop phenology, yield and crop management (including planting date, plant density, fertilization and irrigation) were obtained from the regional trials database of CIMMYT in Zimbabwe. The data from Chisumbanje (19.800 S, 32.867 E), Chiredzi (21.050 S, 31.667 E) and Harare (17.942 S, 31.090 E) stations were used for model calibration while the data from Kadoma (18.369 S, 30.042 E), Makoholi (19.783 S, 30.750 E), Matopos (20.565 S, 28.453 E) and Ratry Arnold Research Station (17.183 S, 31.103 E) were used for model evaluation. Soil profile data of experimental stations were taken from Nyamapfene (1991). Daily rainfall, maximum and minimum temperature and radiation data of the experimental stations were obtained from the respective research stations or nearby meteorological observatories. Estimated data was provided by National Aeronautics and Space Administration-Prediction of Worldwide Energy Resource

(NASA-POWER) (http://power.larc.nasa.gov/) were used whenever radiation data were missing or unavailable.

Model Calibration and Evaluation

The maize model used for the study requires six genetic coefficients which govern the life cycle and reproductive growth of maize cultivars (Table 1). A stepwise iterative calibration procedure was followed whereby genetic coefficients which determine anthesis and physiological maturity dates (P1, P2, and P5) were adjusted in the first stage of the process, followed by those coefficients which affect yield (G2 and G3) using 38 variety-site-year datasets. Rooting profile and soil fertility factors were adjusted with G2 and G3 whenever necessary. Model evaluation was made using an independent dataset (up to 98 variety-site-years). The agreement between simulated and measured values during calibration and evaluation was assessed using root mean square error (RMSE) and index of agreement (d) (Willmott 1982).

Data for Spatial Crop Modeling

The calibrated and evaluated model was then used to simulate the yield of newly-released DT and the commercial check maize varieties in the respective countries at a pixel (≈ 10 km x 10 km) level across the maize growing areas in the study countries (Figure 1).

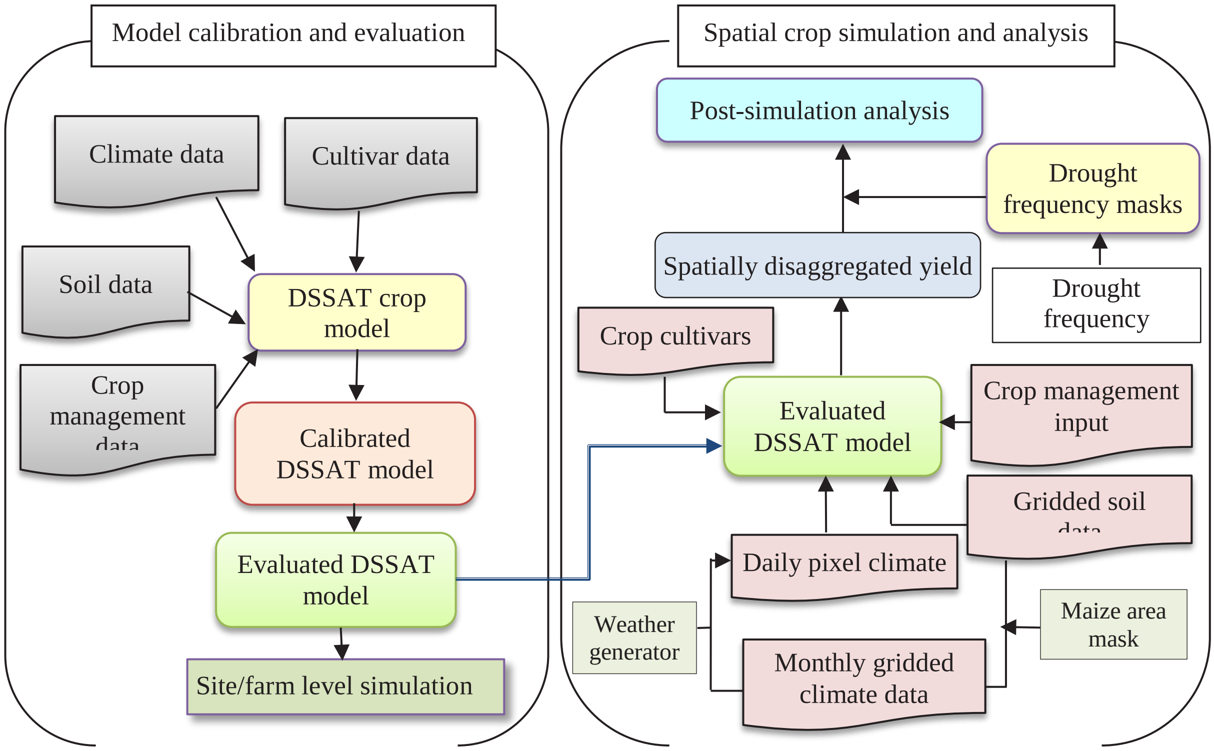


Figure 1. The process followed in crop model calibration, evaluation and spatial simulation.

The spatial simulations were made in a High-Performance Computing cluster (HPC) using gridded climate, soil and crop management data obtained from different online sources. The Spatial Allocation Model (SPAM) raster map for maize (You and Wood 2006) was used to select maize-growing areas in the study countries using the Geographic Resources Analysis

Support System (GRASS) software (http://grass.osgeo.org/). For each grid cell, soil inputs to the model were obtained from a set of twenty–seven generic soil profiles (HC27) developed by blending and interpreting information from both the Harmonized World Soil Database (HWSD) and the World Inventory of Soil Emission (WISE) database based on texture, rooting depth and organic carbon content (Batjes 2009). Simulations were run for all soils in each grid cell, and the cell-specific output was computed from the area-weighted average, based on the area share of each soil in the grid cell. Long-term climate data (1950-2000) for each simulation grid cell were obtained from the Worldclim gridded dataset (Hijmans et al. 2005) which provided all the required climatic elements needed by the stochastic daily weather generator in DSSAT.

A rule-based automatic planting was used to determine area-specific sowing date. The rule refers to a 70% soil moisture within 30-cm soil depth, monthly maximum temperature of <50 ^{o}C and minimum temperature of >7 ^{o}C within a 135-day planting window. The maize varieties were sown at a rate of 5.3 plants m^{-2} and an average of 1000 kg ha^{-1} crop residue was used as initial residue input to the model. All varieties were simulated with two equal split applications of 200 kg ha^{-1} nitrogen. Details on spatial simulation of maize can be found in Tesfaye et al. 2015a.

Evaluation of Variety Performance and Seed Requirement Estimation

The performance of the new DT varieties across the maize growing environments was measured by comparing their yield with the commercial check. Volume of seed required to cover an area of maize with a simulated yield advantage of at least 5% from any of the new DT varieties was determined by multiplying the area by the recent DT maize adoption rate reported for each country using an average seed rate of 25 kg ha^{-1} (CIMMYT 2013b). The seed rate of maize (kg ha^{-1}) varies with the required plant population per hectare, seed weight, seed germination percentage and field loss (Macrobert et al. 2014). In Eastern and Southern Africa, 25 kg ha^{-1} is mostly used as a recommended seed rate for maize (Langyintuo et al. 2008) for a target plant population of approximately 44,000–54,000 plants ha^{-1} depending on the seed weight of varieties (Macrobert et al. 2014).

Results

Drought Frequency

Analysis of drought frequency indicates that all countries in southern Africa are prone to drought during the main cropping season (Figure 2). In the four study countries alone, more than 1.0 million hectares (Mha) of maize growing areas are exposed to seasonal drought events exceeding 20% while an additional 1.6 Mha experience a drought occurrence of 10–20%.

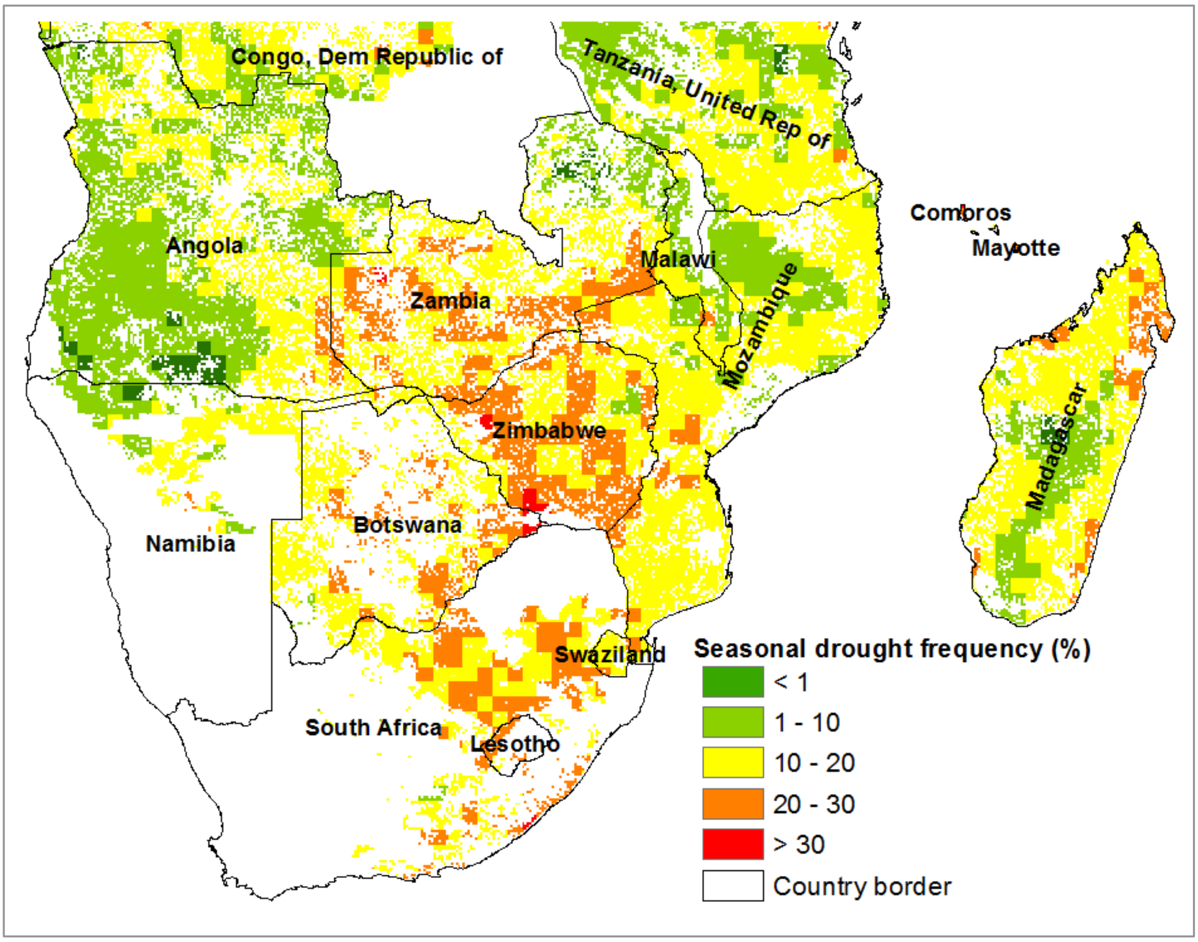


Figure 2. Prevalence of drought in the maize growing areas of southern Africa (1960-1998).

Maize area coverage and frequency of drought vary across MMEs in the study countries. The spatial distribution of maize area and drought frequency across countries and MMEs is presented in Figure 3 while the maize area under different drought frequencies across MMEs is summarized in Figure 4. Most of the maize area is found in the wet upper and wet lower mid-altitude MMEs in Malawi and Zambia, whereas it is located in the dry lowland, wet lowland and wet lower mid-altitude MMEs in Mozambique (Figures 3 and 4). Among the four study countries, Zimbabwe is the only country that has considerable maize area in the dry mid-altitude MME but has no maize area at all in the wet lowland MME. Although the maize area under the highland MMEs is extremely small in all countries, Malawi grows more maize in the highland MME than other countries (Figures 3 and 4). In terms of drought prevalence, Zimbabwe and Zambia are prone to more frequent drought events than that of Malawi and Mozambique across all MMEs (Figure 3). In Zimbabwe, most (>10%) of the seasonal droughts occur in the dry lowland, dry mid-altitude, wet lower mid-altitude and wet upper mid-altitude MMEs comprising a total maize area of 1.2 Mha . In Zambia, most of the maize areas (0.50 Mha) that are exposed to drought occurrences of 20% and above are located in the wet lower mid-altitude and wet upper mid-altitude MMEs (Figure 4). Most of the less frequent seasonal droughts (<15%) occur in the wet lower and wet upper mid-altitude MMEs in Malawi, in the wet lowland and wet lower mid-altitude MMEs in Mozambique and in the wet upper mid-altitude MME in Zambia (Figures 3 and 4).

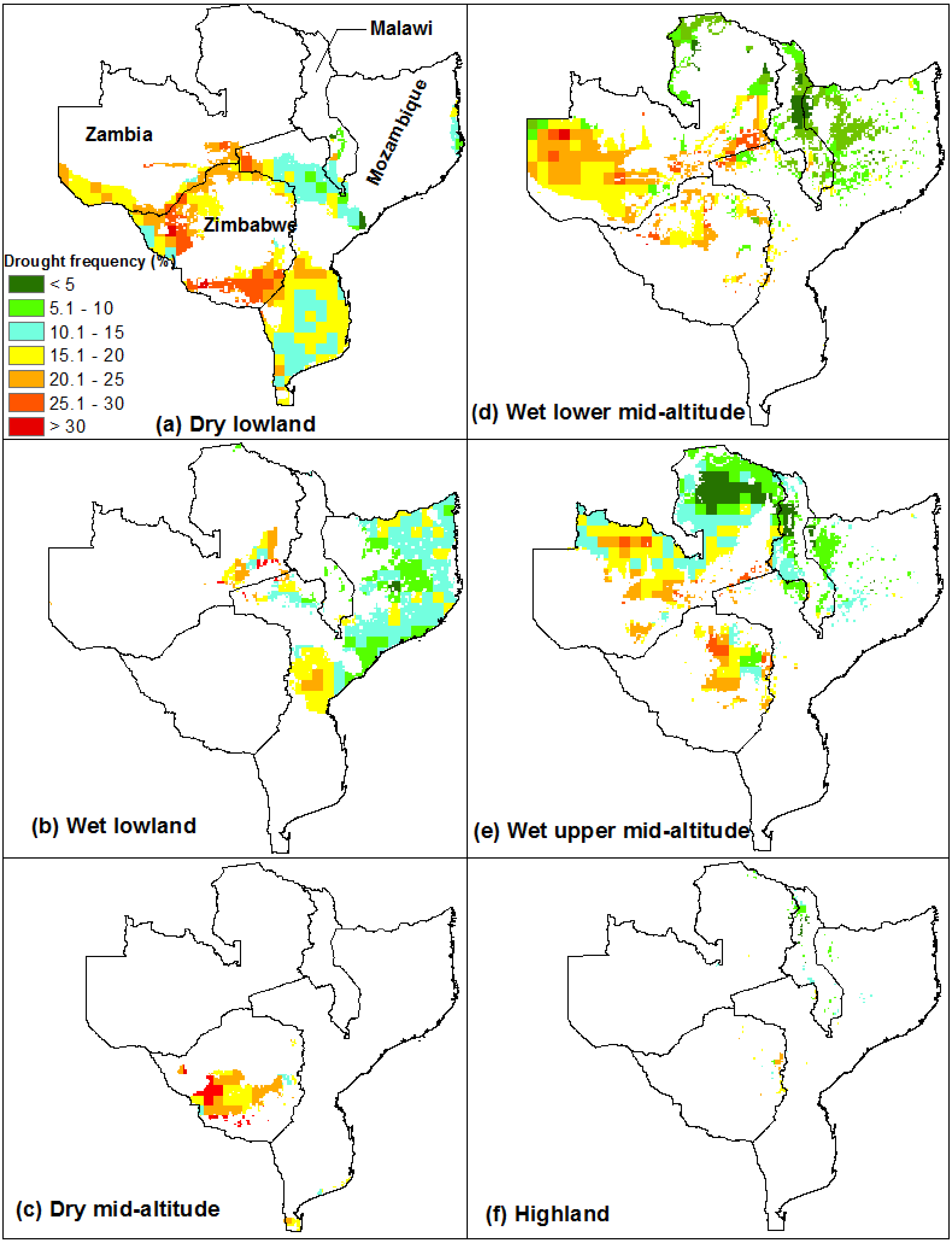


Figure 3. Prevalence of seasonal (November–April) drought (1960–1998) across six maize mega-environment in four southern Africa countries.

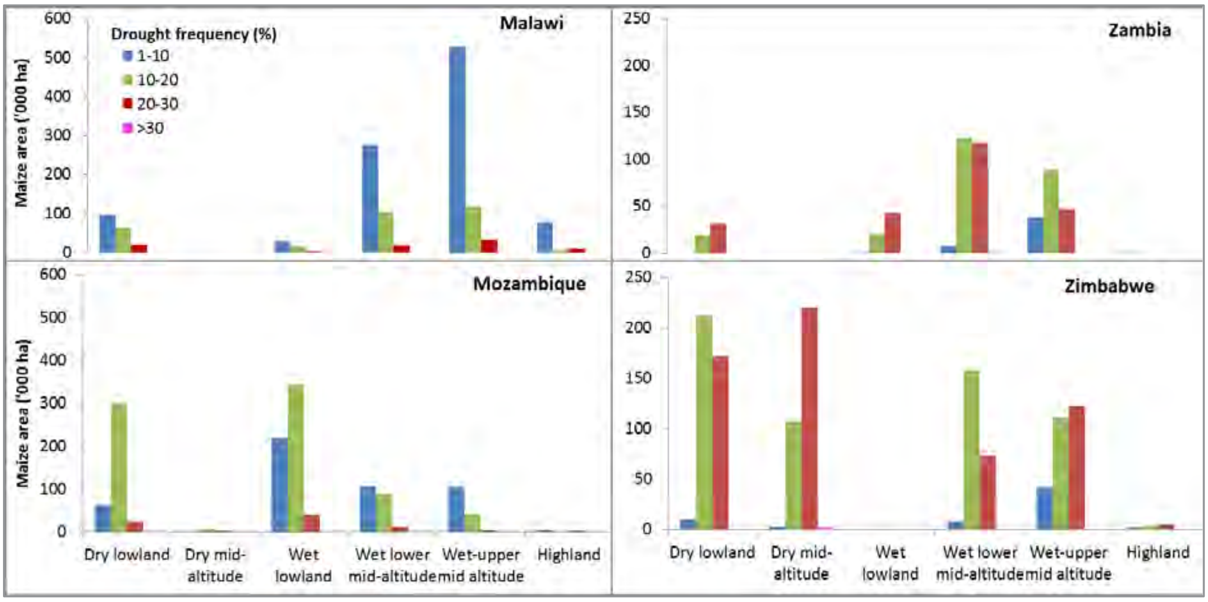


Figure 4. Seasonal drought frequencies across maize mega-environments in four southern Africa countries.

Model Calibration and Evaluation

A comparison of measured and simulated days to anthesis and maturity of the studied maize varieties showed good agreement between the measured and simulated values for both the calibration and evaluation datasets. The average RMSE of days to anthesis and maturity respectively was 4.2 and 7.7 days for the calibration dataset and 3.9 and 2.3 days for the evaluation dataset. The d-index values were 0.94 and 0.74 for days to anthesis and 0.67 and 0.95 for days to physiological maturity in the calibration and evaluation datasets, respectively (see Figure 1 for a plot of measured and simulated values). For grain yield, the average RMSE was 1.6 and 1.0 t ha^{-1} for the calibration and evaluation datasets, respectively. The average simulated yield of the studied varieties across all site-years was closely related to measured grain yield with a d-index of >0.89 both in the calibration and evaluation datasets (see Figure 2 for a plot of measured and simulated grain yield). In general, the indices used for comparing the measured and simulated values of days to anthesis and physiological maturity and grain yield indicate that the CSM–CERES–maize model has captured the response of the DT maize varieties to different growing environments.

Simulated Performance of DT Maize Varieties Across Environments

The simulated relative yield performance of each of the new four DT varieties over that of the standard commercial check is shown in Figure 5. The simulated maize yield across different drought environments indicates that new DT varieties could give a yield advantage of 5% – 40% over the check variety (Figure 6). Although the performance of the new DT varieties varied across environments, they could give an average yield advantage of 16% and 12% under highly (>30% frequency) and less (<10% frequency) drought-prone environments, respectively. Specifically, new DT varieties give 11.4%, 12.9%, 13.6% and 14.7% higher yield than the check across environments with different drought frequencies in Malawi, Mozambique, Zambia and Zimbabwe, respectively. Average yield advantage among new DT varieties ranges from 5%-11% (CZH0946, CZH0811, and CZH0835), 15%-20% (CZH0616) and 28–40% (CZH0837). However, the new DT varieties do not beat the check universally (Figure 5). The coefficient of variation (CV) of yield showed that the new DT varieties could reduce annual yield variability by 3–7% as compared to the commercial check.

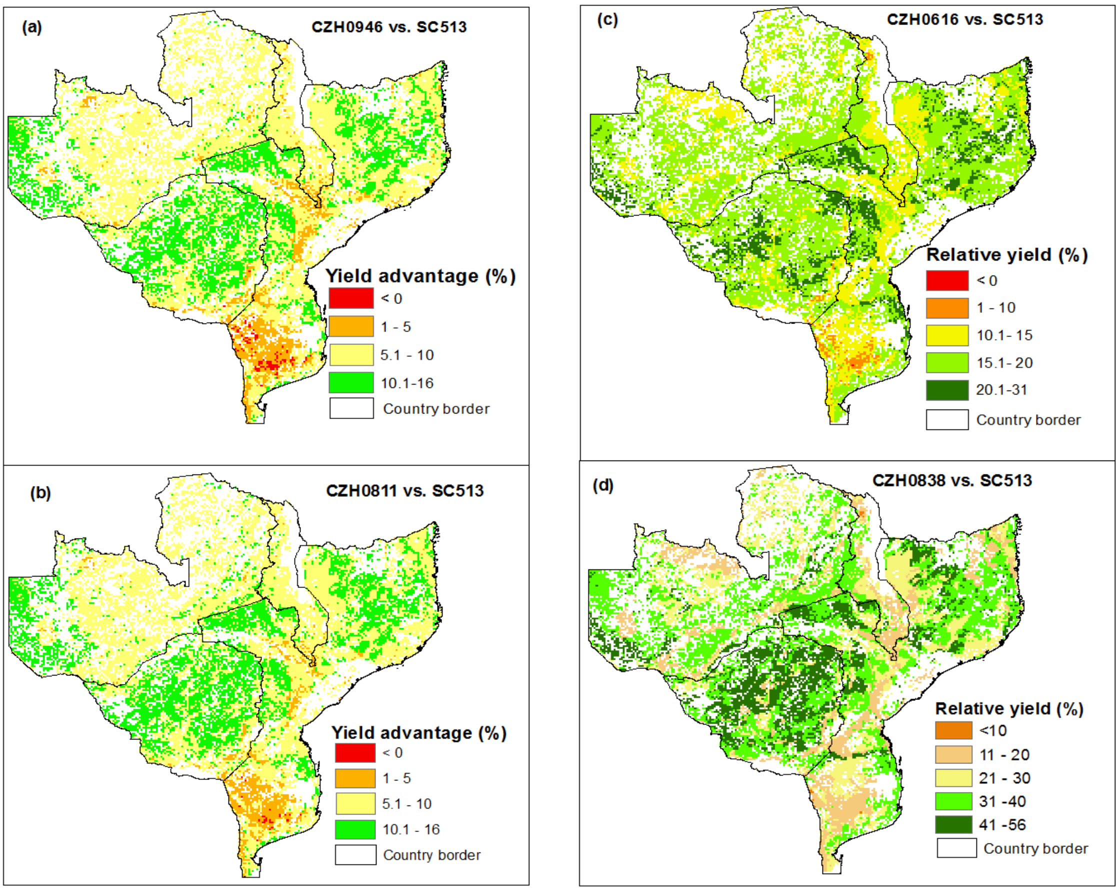


Figure 5. Spatial distribution of simulated relative yields of four new drought-tolerant varieties (a. extra early, b. early, c. medium and d. late maturity) compared to a commercial check (SC513) in four southern Africa countries.

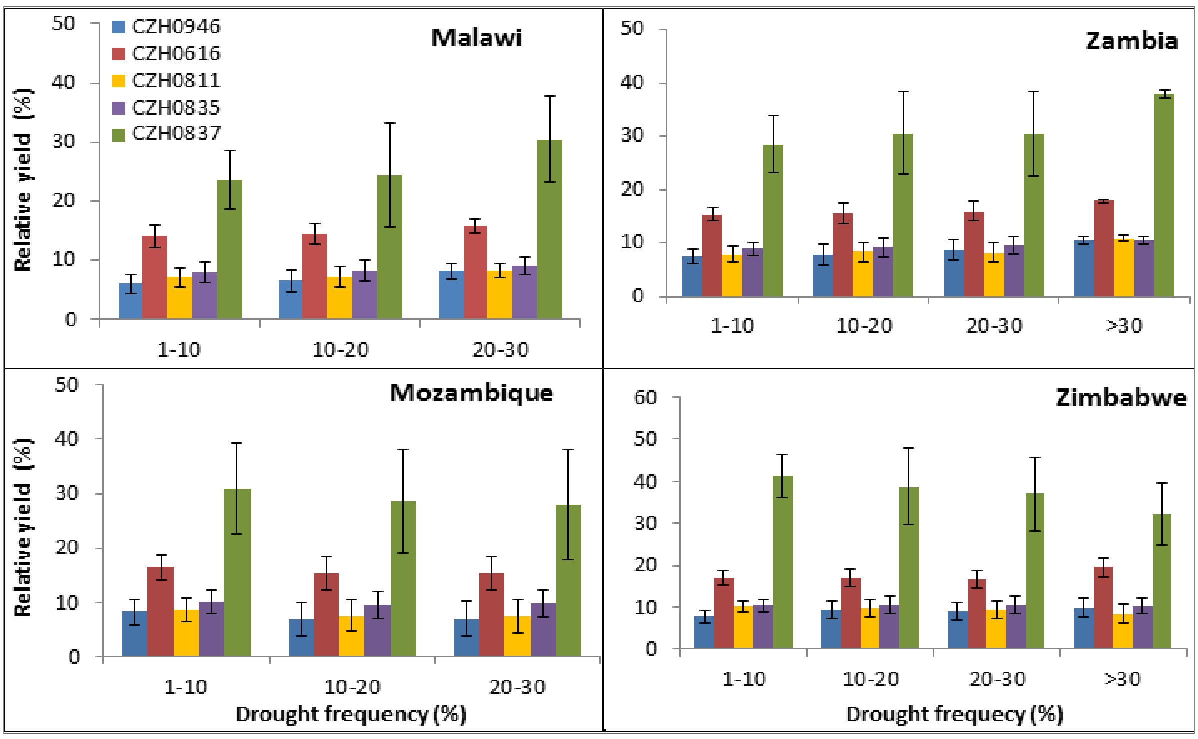


Figure 6. Simulated relative yield advantage and variance of five new drought-tolerant varieties over a commercial check (SC513) across different drought frequency environments in southern Africa. Vertical bars indicate standard deviations.

Potential DT Maize Area and DT Seed Demand

The potential DT maize area and DT seed demand were derived based on the simulated yield advantage (>5%) of the new DT varieties over the commercial check (Table 1). The results show DT maize to have substantial promise in terms of market opportunity for seed companies in the study countries. The level of adoption of new maize varieties varies among countries and so does the potential annual seed requirement: from 5,276 metric tons in Mozambique to 22,302 metric tons in Zimbabwe (Table 1).

Table 1. Potential DT maize area and DT seed demand

Country	Potential DT maize area (ha pa)*	Current DT maize adoption rate (%)**	Potential DT seed demand (metric tons pa)	Current DT seed supply (metric tons pa)***
Malawi	1,387,790	47.3	16,411	4,416
Mozambique	1,366,799	15.4	5,276	855
Zambia	537,092	72.6	9,748	3,422
Zimbabwe	1,251,157	71.3	22,302	7,618

* Based on crop simulation, including all current maize area with a simulated yield advantage of >5% from new DT varieties over commercial check.
** **Source.** CIMMYT (2013b)
*** **Source.** Abate (2013).

Discussion

The highly variable yield of rain-fed crops is the most important downside risk that farmers face in SSA essentially due to the uncertainty surrounding the frequency, intensity, and temporal and spatial distribution of drought (Kassie et al. 2012; Shiferaw et al. 2014). Understanding the nature of drought in a given area is the first step towards managing the risks associated with it (Kassie et al. 2012). Therefore, using long-term gridded data, this study identified the frequency and spatial distribution of seasonal drought during the main cropping season in the major maize growing countries in southern Africa. The results indicated that all the study countries are prone to drought despite variations in drought frequencies. Maize-growing areas in Zambia and Zimbabwe experience more frequent drought events than those in Malawi and Mozambique. The dry lowland and dry mid-altitude MMEs are generally prone to higher drought frequency than the rest of the MMEs, but the size of maize area affected by frequent drought within each MME varies among the study countries. Although all MMEs in Zimbabwe are prone to frequent droughts, the largest drought prone (≥20% frequency) maize area is found in the dry lowland and dry mid-altitude MMEs. In Zambia, however, the largest drought prone maize area is found in the wet lower mid-altitude MME. Therefore, the spatially explicit drought frequency maps generated in this study could be used to design appropriate drought risk management strategies in the respective countries such as targeting DT maize varieties.

Crop models have emerged as potential tools in agricultural research and development and in the exploration of management and policy decisions (Boote et al. 1996), and they have been used to assess spatial and temporal yield variability over different environmental conditions (Batchelor et al. 2002). However, the credibility of outputs of crop models depends on their calibration and evaluation within target environments (Timsina and Humphreys 2006; Xiong et al. 2008). In this study, the CERES–Maize model was calibrated and evaluated for selected DT maize varieties using measured data from a network of maize experiment stations in Zimbabwe. The evaluation results indicate that the model performed well in simulating the phenology and yield of maize after it is calibrated, and results agreed with previous studies that utilized field trial data from different environments to estimate maize genetic coefficients (Gungula et al. 2003; Yang et al. 2009).

This study provided a framework for evaluating the performance of new DT varieties across environments in southern Africa using geospatial analysis and spatial crop modeling tools that allow for an integrated analysis of big datasets (climate, soil, crop and management). Geospatial analysis tools play a valuable role in genotype targeting and can unravel genotype-by-environment interactions by providing high-resolution spatial and temporal data. Spatial analysis is key to identifying environmental frequencies and mapping out target environments that ultimately lead to a more effective deployment of germplasm (Hyman et al. 2013). As shown in this study and previous ones (Hyman et al. 2013; Tesfaye et al. 2015b), spatially explicit crop modeling takes into account changes in year-to-year environmental conditions across environments and could facilitate delivery of the right genotypes to farmers. Since crop varieties or genotypes could perform differently in different environments, a combination of crop simulation models and geographic information systems (GIS) are useful to understand the spatial and temporal aspects of genotype-by-environment interactions (Löffler et al. 2005). In this study, for example, the new DT varieties outperformed the commercial check variety across several environments, but they did not perform better than the check in all environments. Similarly, all new DT varieties did not perform the same way in the same environment, indicating the need for proper targeting of each variety.

Like other modeling studies (e.g., Challinor et al. 2009; Ruane et al. 2013), our study involved some important assumptions. Firstly, except for the varietal change—all other things were assume constant. Given the change of one hybrid seed for another at basically the same seed cost is a common practice in the study region; this appears to be a reasonable assumption. The seed change would not also initially trigger a different crop management practices given the stochastic nature of drought. Over time, however, one would expect farmers to realize the reduced risk inherent in DT maize and possibly adapt maize management practices that potentially increase DT maize benefits further. Secondly, our study focused only on sole maize cultivation and does not simulate other cropping systems such as crop rotation, intercropping or double cropping. Thirdly, the study assumed that plant nutrients other than nitrogen are applied or available in enough quantity so that they do not limit maize growth and development. Our interest in this study is on drought which is more difficult to manage than other crop management practices under rain-fed systems, and hence, our assumptions avoid confounding effects of other factors with drought. This indicates scope for future studies in addressing the assumptions made in this study.

The maps generated in this study show how the new DT varieties perform relative to the commercial check in different environments where maize is currently grown. The results reported in this simulation study are in agreement with previous studies that compared the performance of new DT varieties with commercial checks using field experiments. For example, in less drought prone environments (environments with a yield of ≥3 t/ha), the best DT hybrids yielded 15–25% more than SC513 under on-farm trials in Southern Africa (Setimela et al. 2013). Under severe drought stress environments, DT hybrids gave up to 40% yield advantage compared to commercially available hybrids in the farmers' fields (Setimela et al. 2012; Setimela et al. 2013). Moreover, the field experiments indicated that the best new DT hybrids out-yielded the farmers' own varieties by an average of 35% and 25% under high and low drought conditions in southern Africa, respectively (Setimela et al. 2013). In general, the yield gap between the commercial and the new DT varieties is higher under stressful conditions than non-stressed ones (Bänziger et al. 2006; Edmeades 2013; Setimela et al. 2013) indicating that more progress has been made in developing varieties for drought conditions compared to optimum environmental conditions.

The results of this study also do shed light on the location and volume of potential demand for DT seed and, therefore, could help boost the dissemination of varieties to the farmers that need them. Targeting of new genotypes is not only important to farmers, but it is also critical for public and private seed companies for planning proper marketing and advisory schemes for their varieties (Annicchiarico 2002). The results from this study indicate that the potential annual DT seed volumes in areas where the new DT varieties outperform provide a substantial market opportunity in the four study countries. This helps identify market opportunities for seed companies in southern Africa where varietal replacement is still very slow. However, the potential annual seed volume varies among the countries due to differences in adoption rate; for example, Mozambique has a very large maize area where the new DT varieties could perform well but with relatively low seed requirement. This reiterates that technology adoption is not only dependent on the biophysical suitability of the technology itself but also on socio-economic, political, cultural and institutional factors that may be of equal or greater importance (Notenbaert et al. 2013). Therefore, this type of analysis not only helps seed companies to determine potential annual seed demand in high adoption areas but also to identify areas where adoption is low so that they will be able to plan for addressing the low adoption problems. The relevance of geospatial crop modeling in agribusiness can be further strengthened by integrating socioeconomic factors into the modeling framework (e.g. Tesfaye et al. 2015b).

Conclusion

The availability of big data—soil, climate, elevation and crop distribution—keeps improving over time and there is a growing interest in analytical tools that enable users to handle such data for agricultural applications. This study used geospatial and crop-modeling tools to processes and analyze big datasets for the characterization of drought prevalence and evaluation of the performance of new DT varieties across environments in southern Africa. This type of analysis helps target new DT varieties where they perform well and benefit most and identifies market opportunities. Big data and analytical tools thus can improve the effectiveness of targeting and enhance the uptake of new agricultural technologies that are required in boosting rural livelihoods, agribusiness development and food security in developing countries.

Acknowledgement

We acknowledge a financial support from the Drought Tolerant Maize for Africa (DTMA) project funded by the Bill and Melinda Gates Foundation, USA. We also acknowledge the comments received from anonymous reviewers and the editor which helped improve the quality of the paper. The views expressed here are those of the authors and do not necessarily reflect the views of the authors' institutions or donor. The usual disclaimer applies.

References

Abate, T. 2013. DTMA III Highlights for 2012/13: An Overview. Presented at the Drought Tolerant Maize for Africa (DMA) Annual Meeting. September, Nairobi, Kenya.

Abate, T. B. Shiferaw, A. Menkir, D. Wegary, Y. Kebede, K. Tesfaye, M. Kassie, G. Bogale, B. Tadesse, and T. Keno. 2015. Factors that transformed maize productivity in Ethiopia. *Food Security* 7(5): 965–981.

Annicchiarico, P. 2002. *Genotype x Environment Interactions: Challenges and Opportunities for Plant Breeding and Cultivar Recommendations*. 174. Food and Agriculture Organization, Rome, Italy. http://www.fao.org/docrep/005/y4391e/y4391e00.HTM.

Bänziger, M., P.S. Setimela, D. Hodson, and B. Vivek. 2006. Breeding for improved abiotic stress tolerance in maize adapted to southern Africa. *Agricultural Water Management* 80(1): 212–224.

Batchelor, W. D., B. Basso, and J. O. Paz. 2002. Examples of strategies to analyze spatial and temporal yield variability using crop models. *European Journal of Agronomy* 18:141–158.

Batjes, N. H. 2009. Harmonized soil profile data for applications at global and continental scales: updates to the WISE database. *Soil Use and Management* 25(2):124–127.

Bezu, S., G. T. Kassie, B. Shiferaw, and J. Ricker-Gilbert. 2014. Impact of improved maize adoption on welfare of farm households in Malawi: A panel data analysis. *World Development* 59(C): 120–131.

Boote, K. J., J. W. Jones, and N. B. Pickering. 1996. Potential uses and imitations of crop models. *Agronomy Journal* 88: 704–716. doi:10.2134/agronj1996.00021962008800050005x.

Cairns, J.E., J. Crossa, P.H. Zaidi, P. Grudloyma, C. Sanchez, J.L. Araus, S. Thaitad, D. Makumbi, C. Magorokosho, M. Bänziger, A. Menkir, S. Hearne, and G.N. Atlin. 2013. Identification of drought, heat, and combined drought and heat tolerant donors in maize. *Crop Science* 53(4): 1335–1346.

Challinor, A.J., F. Ewert, S. Arnold, E. Simelton, and E. Fraser. 2009. Crops and climate change: progress, trends, and challenges in simulating impacts and informing adaptation. *Journal of Experimental Botany* 60(10): 2775–89.

CIMMYT. 2013a. The Drought Tolerant Maize for Africa project. DTMA Brief, September. http://dtma.cimmyt.org/index.php/about/background [accessed October 14, 2015].

CIMMYT. 2013b. Country level improved maize adoption monitoring surveys. International Maize and Wheat Improvement Center, Mexico, D.F.

Edmeades, G.O. 2013. Progress in Achieving and Delivering Drought Tolerance in Maize - An Update. *International Service for the Acquisition of Agri-biotech Applications* 1–44. http://www.isaaa.org [accessed May14, 2016].

Erenstein, O., J. Hellin, and P. Chandna. 2010. Poverty mapping based on livelihood assets: A meso-level application in the Indo-Gangetic Plains, India. *Applied Geography* 30(1): 112–125.

FAOSTAT. 2015. FAO statistical databases. Food and Agriculture Organization of the United Nations. Statistics Division. Rome, Italy. http://faostat3.fao.org/home/E.

Fisher, M., T. Abate, R. W. Lunduka, W. Asnake, Y. Alemayehu, and R. B. Madulu. 2015. Drought tolerant maize for farmer adaptation to drought in sub-Saharan Africa: Determinants of adoption in eastern and southern Africa. *Climatic Change* 133(2): 283–299.

Gungula, D. T., J. G. Kling, and A. O. Togun. 2003. CERES-maize predictions of maize phenology under nitrogen-stressed conditions in Nigeria. *Agronomy Journal* 95(4): 892–899. doi:10.2134/agronj2003.8920.

Hijmans, R. J., S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis. 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* 25(15): 1965–1978. doi:10.1002/joc.1276.

Hodson, D. P., E. Martínez-Romero, J. W. White, J. D. Corbett, and M. Bänziger. 2002. Maize Research Atlas, version 3.0. Available from http://www.cimmyt.org/en/about-us/job-opportunities/1441-k12255-germplasm-data-coordinator-kenya [accessed October 15, 2015].

Homann-Kee Tui, S., M. Blümmel, D. Valbuena, A. Chirima, P. Masikati, A.F. van Rooyen, and G.T. Kassie. 2013. Assessing the potential of dual-purpose maize in southern Africa: A multi-level approach. *Field Crops Research* 153: 37–51.

Hoogenboom, G. et al. 2010. Decision Support System for Agrotechnology Transfer version 4.5 [CD-ROM]. University of Hawaii, Honolulu, HI.

Hyman, G., D. Hodson, and P. Jones. 2013. Spatial analysis to support geographic targeting of genotypes to environments. *Frontiers in Physiology* 4: 1-13.

IRI. 2015. Dataset: IRI Analyses SPI: Standardized Precipitation Index analyses of multiple global precipitation datasets. http://iridl.ldeo.columbia.edu/SOURCES/.IRI/.Analyses/.SPI/SOURCES/.IRI/.Analyses/.SPI/ [accessed April 19, 2015].

Jones, C. A., and J. R. Kiniry. 1986. CERES-Maize: A simulation model of maize growth and development. In *CERES Maize a Simulation Model of Maize Growth and Development,* edited by C. A. Jones and J. R. Kiniry. Texas A&M University Press, Texas.

Jones, J. W., G. Hoogenboom, C. H. Porter, K. J. Boote, W. D. Batchelor, L. A. Hunt, P. W. Wilkens, U. Singh, A. J. Gijsman, and J. T. Ritchie. 2003. The DSSAT cropping system model. *European Journal of Agronomy* 18: 235–265.

Kassie, G. T., O. Erenstein, W. Mwangi, R. Larovere, P. Setimela, and A. Langyintuo. 2012. Characterization of Maize Production in Southern Africa : Synthesis of CIMMYT / DTMA Household Level Farming System Surveys. 4. Mexico, D.F.

Kassie, G. T., R. La Rovere, W. Mwangi, O. Erenstein, A. Langyintuo, and K. Sonder. 2013. Drought Risk and Maize Production in Southern Africa. *Journal of Asian Scientific Research* 3: 956–973.

Kassie, M., M. Jaleta, and A. Mattei. 2014. Evaluating the impact of improved maize varieties on food security in Rural Tanzania: Evidence from a continuous treatment approach. *Food Security* 6: 217–230.

Lang, C., C.B. Barrett, and F. Naschold. 2013. Targeting Maps: An Asset-Based Approach to Geographic Targeting. *World Development* 41(1):232–244.

Langyintuo, A.S., A.O. Diallo, J.F. MacRobert, J. Dixon, and M. Bazinger. 2008. An analysis of the bottlenecks affecting the production and deployment of maize seed in Eastern and Southern Africa. International Maize and Wheat Improvement Center (CIMMYT), Mexico, D.F.: CIMMYT.

Löffler, C. M., J. Wei, T. Fast, J. Gogerty, S. Langton, M. Bergman, B. Merrill, and M. Cooper. 2005. Classification of maize environments using crop simulation and geographic information systems. *Crop Science* 45(5):1708. doi:10.2135/cropsci2004. 0370.

Macrobert, J.F., P. Setimela, J. Gethi, and M.W. Regasa. 2014. Maize Hybrid Seed Production Manual. International Maize and Wheat Improvement Center (CIMMYT), Mexico, D.F.: CIMMYT.

Mckee, T. B., N. J. Doesken, and J. Kleist. 1993. The relationship of drought frequency and duration to time scales: Proceedings of the 8th Conference on Applied Climatology. 17(22) California.

Mitchell, T. D., and P. D. Jones. 2005. An improved method of constructing a database of monthly climate observations and associated high-resolution grids. *International Journal of Climatology* 25:693–712.

Notenbaert, A., M. Herrero, H. De Groote, L. You, E. Gonzalez-Estrada, and M. Blummel. 2013. Identifying recommendation domains for targeting dual-purpose maize-based interventions in crop-livestock systems in East Africa. *Land Use Policy 3*0: 834–846.

Nyamapfene, K. W. 1991. *The Soils of Zimbabwe*. Nehanda Publisher, Harare.

Ray, D. K., N. Ramankutty, N. D. Mueller, P. C. West, and J. A. Foley. 2012. Recent patterns of crop yield growth and stagnation. *Nature Communications* 3: 1293. doi: 10.1038/ ncomms2296.

Rockström, J., and J. Barron. 2007. Water productivity in rainfed systems: overview of challenges and analysis of opportunities in water scarcity prone savannahs. *Irrigation Science* 25(3):299–311.

Ruane, A.C., L.D. Cecil, R.M. Horton, R. Gordón, R. McCollum, D. Brown, B. Killough, R. Goldberg, A.P. Greeley, and C. Rosenzweig. 2013. Climate change impact uncertainties for maize in Panama: Farm information, climate projections, and yield sensitivities. *Agricultural and Forest Meteorology* 170(March): 132–145.

Sabarina, K., and N. Priya. 2015. Lowering data dimensionality in big data for the benefit of precision agriculture. *Procedia Computer Science* 48: 548–554.

Shiferaw, B., B. M. Prasanna, J. Hellin, and M. Bänziger. 2011. Crops that feed the world 6. Past successes and future challenges to the role played by maize in global food security. *Food Security* 3: 307–327.

Shiferaw, B., K. Tesfaye, M. Kassie, T. Abate, B. M. Prasanna, and A. Menkir. 2014. Managing vulnerability to drought and enhancing livelihood resilience in sub-Saharan Africa: Technological, institutional and policy options. *Weather and Climate Extremes* 3: 67–79.

Setimela, P., J. MacRobert, G. N. Atlin, C. Magorokosho, A. Tarekegne, D. Makumbi and G. Taye. 2012. Performance of elite maize varieties tested on-farm trials in eastern and southern Africa. Presented at the International Annual Meeting, American Society of Agronomy/ Crop Science Society of America/ Soil Science Society of America Meetings. October. Cincinnati, Ohio.

Setimela, P.S., G. Tesfahun, O. Erenstein, and O.F. Ndoro. 2013. Performance of Elite Drought Tolerant Maize Varieties Tested On-Farm in Eastern and Southern Africa. Presented at the International Annual Meeting, American Society of Agronomy/ Crop Science Society of America/ Soil Science Society of America Meetings. November. Tampa, Florida.

Sienz, F., O. Bothe, and K. Fraedrich. 2012. Monitoring and quantifying future climate projections of dryness and wetness extremes: SPI bias. *Hydrology and Earth System Sciences* 16: 2143–2157.

Tesfaye, K., S. Gbegbelegbe, J. E. Cairns, B. Shiferaw, B. M. Prasanna, K. Sonder, K. Boote, D. Makumbi, and R. Robertson. 2015a. Maize systems under climate change in sub-Saharan Africa. *International Journal of Climate Change Strategies and Management* 7(3): 247–271.

Tesfaye, K., M. Jaleta, P. Jena, and M. Mutenje. 2015b. Identifying potential recommendation domains for conservation agriculture in Ethiopia, Kenya, and Malawi. *Environmental Management* 55(2): 330-46. doi: 10.1007/s00267-014-0386-8.

Timsina, J., and E. Humphreys. 2006. Performance of CERES-Rice and CERES-Wheat models in rice–wheat systems: A review. *Agricultural Systems* 90:5–31.

Wani, S. P., T. K. Sreedevi, J. Rockstroma, and Y. S. Ramakrishna. 2009. Rainfed Agriculture – Past Trends and Future Prospects. In *Rainfed Agriculture: Unlocking the Potential*. 1–35. http://oar.icrisat.org/2428/ [accessed October 15, 2015].

Weber, V. S., A. E. Melchinger, C. Magorokosho, D. Makumbi, M. Bänziger, and G. N. Atlin. 2012. Efficiency of managed-stress screening of elite maize hybrids under drought and low nitrogen for yield under rainfed conditions in Southern Africa. *Crop Science* 52:1011. doi: 10.2135/cropsci2011.09.0486.

Xiong, W., I. Holman, D. Conway, E. Lin, and Y. Li. 2008. A crop model cross calibration for use in regional climate impacts studies. *Ecological Modelling* 213(3-4):365–380.

Yang, Z., G. G. Wilkerson, G. S. Buol, D. T. Bowman, and R. W. Heiniger. 2009. Estimating genetic coefficients for the CSM-CERES-Maize model in North Carolina environments. *Agronomy Journal* 101(5):1276-1285.

You, L., and S. R. Wood. 2006. An entropy approach to spatial disaggregation of agricultural production. *Agricultural Systems* 90:329–347. doi:10.1016/j.agsy.2006.01.008.

International Food and Agribusiness Management Review
Special Issue - Volume 19 Issue A, 2016

Index Insurance: Using Public Data to Benefit Small-Scale Agriculture

Maria Jose Castillo[ⓘa], Stephen Boucher[b], and Michael Carter[c]

[a] *Professor, ESPAE Graduate School of Management, Escuela Superior Politécnica del Litoral, Ecuador Malecón 100 y Loja, Guayaquil, Ecuador*

[b] *Associate Professor,Agricultural and Resource Economics Department University of California, Davis and the Giannini Foundation of Agricultural Economics, One Shields Ave., CA 95616, USA*

[c] *Professor and Director of I4, Index Insurance Innovation Initiative, Agricultural and Resource Economics Department, University of California, Davis and the Giannini Foundation of Agricultural Economics, One Shields Ave., CA 95616, USA*

Abstract

This paper highlights the importance of public data for the development of more efficient and sustainable risk management schemes, such as index insurance, for smallholder agriculture. Three case studies of index insurance—catastrophic weather insurance in Mexico, satellite-based insurance for pastoralists in Kenya, and a hypothetical area-yield insurance scheme in Ecuador—are briefly analyzed in terms of the data and type of index used, the way the contract was designed and implemented (or simulated) and the impacts of the insurance on investment, nutrition and income smoothing. The increasing opportunity to use *big data* for improving and expanding index insurance is also addressed. The analysis suggests that the strong potential for index insurance to improve the welfare of small farmers represents a clear justification for increased government investment in the collection of the types of data that can facilitate the expansion of index insurance markets.

Keywords: Index insurance, public data, big data, small farmers, developing countries

[ⓘ]Corresponding author: Tel: + 593.4.208.1084 ext. 131
Email: M. J. Castillo: mjcasti@espol.edu.ec
S. Boucher: boucher@primal.ucdavis.edu
M. Carter: mrcarter@ucdavis.edu

Introduction

Many governments in developing countries have, in the new millennium, prioritized the creation and strengthening of the agricultural crop insurance markets. Multiple motivations underlie these insurance initiatives. First, governments are concerned about the pressure placed on public budgets by the increasing frequency of catastrophic weather events, such as severe droughts, flooding and frost, associated with climate change. Strengthening formal insurance is seen as a more efficient means of managing this risk than ex-post disaster relief. Second, addressing missing or incomplete insurance markets is increasingly viewed as a necessary step to enhance food security and reduce poverty. Without insurance, farm households, especially small-holders who continue to account for the majority of basic grain production in many developing countries, are often unable to make investments because banks refuse to offer credit to uninsured farmers. Those who do have access to credit are often unwilling to seek credit because the collateral requirements would expose them to too much risk (Boucher et al. 2008). The end result of under-developed insurance markets is a vicious cycle of under-investment, stagnant agricultural yields and the persistence of rural poverty.

While the logic of strengthening agricultural insurance markets is clear, the path forward is much less so. One option is to build a market for conventional named peril (or indemnity) insurance. The challenges of creating a broad and sustainable market for conventional insurance contracts in developing country settings, however, are considerable. The combination of information asymmetries and poor infrastructure present the largest hurdle. Conventional contracts require multiple field inspections in order to evaluate losses and determine if they were caused by insurable events instead of farmer negligence (moral hazard). When telecommunications and road infrastructure are poor, the costs of effectively carrying out these types of inspections and overcoming information asymmetries between the farmer and the insurance company can be prohibitively high, thus undermining the viability of the insurance market, unless that it benefits from massive subsidies.[1]

Index insurance represents an attractive alternative, especially in small farmer contexts.[2] Under an index insurance contract, indemnity payments are triggered when an external index, such as a rainfall during the planting season or the average yield of a specific area exceeds (or falls below) a critical value called the strike-point. Since payouts do not depend on the loss experienced by the individual insured farmer, index insurance is less susceptible to asymmetries of information. Similarly, since determining whether a payout is warranted does not require on-farm inspections, index insurance may be offered with substantially lower transaction costs. Against these advantages stands one of the primary challenges of index insurance; namely "basis risk", or the risk that a farmer suffers a loss but does not receive an insurance payout. As described by Carter (2012), some basis risk is unavoidable in index insurance, but it can be minimized by careful contract design that maximizes the correlation between the index and farmers' losses. A number of recent studies analyze the potential of index insurance to reduce poverty by enhancing households' capacity to smooth consumption in the face of weather shocks and improve both households' access to and willingness to take

[1] Skees et al. (2006) offer a detailed description of the costs and challenges of conventional contracts (multi-peril) associated to the lack and asymmetry of information.
[2] See Hazell et al. (2010) for a detailed summary of the evolution of index insurance in developing countries. Barnett et al. (2008) and Carter et al. (2014) present a summary of the pilots of index insurance in the third world.

on credit (Skees 2006; Barnett et al. 2008; Hazell et al. 2010). These authors are careful to point out, however, that the success of index-based insurance will ultimately depend on whether or not the contracts that are designed offer a significant reduction in transaction costs without prohibitively high levels of basis risk.[3]

Although high quality, empirical evidence on the impacts of index insurance is scarce, initial efforts give reasons for cautious optimism. For example, Fuchs and Wolff (2011a) find a statistically significant impact of county-level weather index insurance on maize productivity and household expenditure and income in Mexico. Elabed et al. (2014) find a positive effect of an area-yield index contract on area planted and seed investment among cotton farmers in Mali. Karlan et al. (2012) find that a rainfall-based index insurance contract has stronger effects than direct cash grants on farmers' investment levels in Ghana. Finally, Janzen and Carter (2013) show that a satellite-based index contract that measures the level of natural pasture available reduced distress sales of livestock and significantly stabilized consumption among herders in northern Kenya.

A separate strand of the literature has directly compared index versus conventional insurance in order to identify circumstances under which index insurance has the potential to perform better than conventional insurance. Miranda (1991) theoretically establishes conditions under which area yield-based index insurance would provide greater protection to farmers than conventional insurance. Through a simulation based on Kentucky soybean producers, he then concludes that, for most producers, area-yield index insurance would provide "better overall yield risk protection than individual insurance (p. 242)" since it would cover more of the systemic yield risk and it would be more sustainable than individual insurance. Similarly, Breustedt et al. (2008) find that area-yield insurance is more effective in risk reduction for wheat producers in Kazakhstan than individual-based yield insurance with a low strike yield. Research in Ecuador that will be discussed in this article also reflects the potential of area yield-based index insurance to provide farmers greater risk management than conventional insurance.

Nonetheless, despite its clear advantages, index insurance also faces important challenges; mainly the lack of information required to build an effective index that offers real protection for farmers. Binswanger-Mkhize (2012) questions the general availability of sufficient information and provides a cautionary critique of the recent shift towards index-based insurance. Binswanger-Mkhize's cautionary message is important, as there is no guarantee that sufficient quantity and quality of data will be available to design contracts of sufficiently high quality. The lack of high quality data aggravates basis risk by increasing the frequency and size of index prediction errors and reducing the correlation between farm yields and the index (Carter 2012).[4] Information in developing countries on both yields and potential weather-based indices, such as rainfall and temperatures, are often characterized by low quality, with high frequencies of missing data and short time series. This leads to a vicious cycle; data are not used for productive or effective economic purposes given its low quality, and the lack of demonstrated, valuable uses of the data discourages investment in improved data collection.

[3] Carter (2012) define basis risk as "deviations in yield experienced by the household that are not correlated by deviations in the index and that are therefore uninsured by the index insurance contract (p. 4)."
[4] Other sources of basis risk include idiosyncratic risk and inadequate choices of the geographic scale for the index (See Carter, 2012).

In light of the potential for index insurance to improve small farmers' risk management capacity, we propose a dynamic strategy in which existing data is evaluated as a starting point for getting index insurance markets off the ground. If the index contracts meet a minimum quality threshold, pilot programs could be implemented, with the expectation that additional investment would be made in expanded and improved data collection so that contract quality would improve as the quality and quantity of data grows. Given the high premium rates that insurance and reinsurance companies tend to charge in order to offer index insurance in sparse-data environments, Carter (2013) proposes a public-private reinsurance partnership in which the public sector initially provides some lower-cost reinsurance for index insurance policies. Primary responsibility for reinsurance would then pass to the private reinsurance sector as additional data is accumulated, "parameter uncertainty" is reduced and more affordable contract pricing becomes possible. Especially during the initial pilot phase when basis risk may be high, hybrid contracts that combine a weather, satellite or area-yield-based index with on-farm "audits" represent a promising strategy to compensate for initially high basis risk (Carter et al. 2014).

A primary role of government under this type of arrangement would take the form of increased investment in data collection efforts such as yield surveys and automated weather stations, and potentially investment in technologies that allow the gathering and analysis of higher volumes and variety of information (e.g. GPS, drones, hardware and software appropriate for big data) for agricultural insurance purposes. This strategy for the use of public funds to break the aforementioned vicious cycle is based on the public good nature of this type of yield and weather information and, we expect, promises to be more cost-effective and sustainable than large-scale, direct premium subsidies.

In this paper, we emphasize the potential for large-scale, government collected and/or publicly available data to effectively enhance insurance availability and risk management for smallholder farmers by enabling the design and implementation of index insurance contracts. We do this by summarizing three case studies that utilize three different types of indices, and thus require different types of data. The first case study is Mexico's experience with catastrophic, weather-based index insurance. The second is the Index-Based Livestock Insurance (IBLI) in Kenya, a contract designed with a combination of livestock mortality data and satellite data that measures the vegetative density and thus caloric availability to animals of natural pasture. The third case study presents results of a research project in Ecuador where the potential performance of a hypothetical area yield-based index insurance contract was compared to the actual performance of an existing conventional insurance contract. The potential for and challenges facing the use of big data to improve the quality of index insurance contracts and extend index insurance markets in the future is also addressed in this paper.

The paper is structured as follows. Section two briefly describes the different types of indices that are most commonly used and the advantages and disadvantages of each. The three case studies are presented in Sections three, four and five. Section six links the potentials for big data application to index insurance in developing countries, including examples in the context of the researched cases. The final section concludes.

Types of Index Insurance

The primary objective of index insurance is to protect farmers against covariate risk, or risk that drives fluctuations in average yield of farmers in a given region. The ideal index insurance contract would thus be perfectly correlated with average, or area, yields in the contract area. Based on this goal, we can classify indices into two general classes: *indirect and direct*. Indirect indices use data from weather stations or satellites to generate indirect estimates of average yields in the contract area. Examples of indirect indices include various functions of weather phenomena, such as cumulative rainfall during planting season, and the Normalized Difference Vegetative Index (NDVI), which uses satellite images to estimate the density of pasture available to livestock. An important challenge of indirect indices is understanding the relationship between the weather event (or satellite imagery) that generates the data (i.e., millimeters of rainfall) and average yield and then to design the index to best capture this relationship. In many cases, this requires a good agronomic model of crop growth for the specific insured crops.

The potentially large advantage of indirect indices is the relatively low cost of index measurement which, in many cases, simply requires taking measurements from weather stations or downloading publically available satellite data from the internet. However, in practice, acquiring and assembling data underlying indirect indices may imply some costs. First, there may exist fixed costs to design the index (including research to identify the strongest relationship between the available weather or satellite data and yields). Second, the information may not be freely available. Although it is typically the public sector that collects and manages weather data, the institutions that manage the data may charge for their access. In the case of satellite data, experts often need to be hired to convert the raw data into a form that is usable for the purpose of an index. Finally, installing and maintaining weather stations implies a non-negligible cost.

An important disadvantage of indirect indices is that, if the index only captures one of the multiple sources of covariate risk, then basis risk may be significant. For example, coffee production is adversely affected by excess rainfall in the flowering period as well as by a deficit of solar radiation during the period of fruit growth. If the index is based solely on rainfall, for example, the contract will likely suffer from significant basis risk.

Direct indices, in contrast, directly estimate average, or area, yield in the contract area, typically through a production survey or plant cuttings of randomly selected plots.[5] Precisely because they directly measure average yields, direct indices take into account all of the potential sources of covariate risk that affect average production levels and, as a result, will be characterized by lower levels of basis risk than indirect, weather-based indices. A second advantage of direct indices is that they are typically more intuitive, transparent and easy to understand for farmers compared to indirect indices.

The main disadvantage of direct indices is the greater cost associated with directly measuring average yields through farmer surveys or crop cuttings. This cost will depend on various factors, including the sample size needed to achieve a specified level of statistical precision of the average yield estimate as well as the spatial dispersion of and ease of access to the sampled plots. Another important factor affecting the cost of direct indices is the existence (or not) of a national agricultural production survey upon which area yields can be estimated at a sufficiently disaggregated scale.

Breustedt et al. (2008) show for Kazakhstan the ability of area yield insurance to provide more risk reduction than weather-based index insurance. The benefits of area yield index

[5] In the case of cattle, livestock mortality measured via survey is an example of a direct index.

insurance are also noticed by Carter et al. (2007), who compare the performance of area yield and weather based index insurance for farmers and lenders in Peru.

Catastrophic Weather-Based Index Insurance: The Mexican Case

Mexico is among the countries with the most advanced agricultural insurance programs in the world, including several different types of index insurance programs since 2003.[6] Mexico was one of the first countries to implement a catastrophic weather-based index insurance program (WII). This program has grown rapidly and is now one of the largest worldwide (Fuchs and Wolff 2011a). The contract uses a rainfall index to protect small farmers growing maize, sorghum, beans and wheat, against droughts, the main cause of agricultural catastrophes in Mexico (AGROASEMEX 2006). Below we describe the data used for the WII contract, details about the contract's implementation and its effect on farmer behavior and poverty reduction.

The Data

Mexico's WII contract uses publicly available rainfall and temperature data from the government's network of weather stations. These weather data, along with data on soil types from detailed soil maps, are fed into a dynamic crop model that allows estimation of the relationship between yields and the specific weather phenomenon.[7] The model allows AGROASEMEX, the Mexico's parastatal insurance and re-insurance company to estimate crop yields in regular circumstances and yields when a deficit of precipitation is the primary limiting factor.[8] Thresholds are established for each stage of the crop's vegetative cycle such that when rainfall is below the threshold level, farmers are highly likely to suffer significant yield losses (AGROASEMEX 2006). WII also takes into consideration critical temperature levels that indicate severe loss (AGROASEMEX 2015).

Although the Mexican government manages over 5,000 weather stations, relatively few of them are suitable for WII (Fuchs and Wolff 2011a).[9] Many stations are ruled out because they are not located close enough to areas where the insurable crops are grown. In addition, AGROASEMEX and its international reinsurers require that all weather stations used to develop the index comply with international quality standards. Specifically, the data from a station must be available for at least twenty-five continuous years, with a maximum of 10% missing or invalid data. In addition, the stations must allow timely reading of the climatic data so that contract implementation and potential payouts to farmers are not delayed (AGROASEMEX 2006).[10]

[6] Other types of index insurance implemented by AGROASEMEX include NDVI and area yield index insurance.

[7] By isolating climatic events from other factors that affect production, the model performs simulations that allow the calculation of dry matter under both potential and limiting climatic conditions. The main components of the model are the physiological age of the crop, the raw assimilation of CO2 and dry matter distribution (AGROASEMEX 2006).

[8] Since 2013 a private insurance company is also offering WII together with AGROASEMEX (FAO, 2014).

[9] By 2006, 297 weather stations were participating in Mexico's WII (AGROASEMEX, 2006).

[10] Historical data is used to estimate the probability distribution function of the index, which is crucial for designing the contract. Periodic data is required for the operation of the insurance contract; that is, to determine whether or not a payment is due, and the amount of the covered loss.

As acknowledged by AGROASEMEX (2006), expansion of the area eligible for catastrophic insurance coverage is limited by the lack of data of sufficient quantity and quality. To this end, this company has been working on improving the quality of data generated by the weather stations. This effort is expected to lead to premium reductions by the international re-insurance companies.

Implementation of the Contract

The general goals of the WII contract are to protect low-income farmers from severe climatic shocks and help state and local governments to more efficiently manage the risk of catastrophic losses among the rural population (AGROASEMEX 2006). The premium is fully subsidized with the federal government assuming 80% of the premium cost while state governments pay the remaining 20%. For poorer states the arrangement is 90% (federal) – 10% (state). As a result of the WII program, the federal government has been able to reduce post-catastrophe direct payments to state and local governments; federal government participation fell from 70% to 50% of all direct payments to farmers in 2013 (FAO 2014).

The program started in 2003, with a pilot offering in the state of Guanajuato. Based on its initial success, the program has expanded to vulnerable areas of all thirty-two Mexican states (FAO 2014). The area covered by the program grew remarkably fast, from around 100,000 hectares in 2003 to 12 million hectares in 2013. Total indemnity payments to farmers between 2003 and 2013 exceeded USD 290 million (Ibid).

Each year, states propose to the federal government the specific counties and the number of hectares within each county to be insured before the beginning of the planting season (January to March).[11] When a catastrophic event occurs (rainfall or temperature levels surpassing maximum thresholds or falling below minimum thresholds) indemnity payments are received by the state governments and then distributed to farmers that meet established eligibility criteria (in general, less than 20 hectares) in the insured regions.[12] The beneficiaries are identified only when indemnity payments are due (FAO 2014).

In order for the WII program to have its intended consequences, for example of increasing investment levels by small-holders, eligible farmers in the insured areas should know that they are insured and should receive support in case of a climatic shock. It is also important that the farmers understand that any indemnity payments they receive are the results of an established insurance market instead of political or other types of interests. In order to promote awareness of coverage among eligible farmers, the government seeks to inform them about the insurance through regional offices of official programs such as PROCAMPO. In addition, the Ministry of Agriculture contracts external evaluations of the program and includes among program indicators farmers' familiarity with the insurance. These outreach efforts appear quite successful; according to the 2010 evaluation (Universidad Autónoma Chapingo 2010), 95.5% of the covered population knew about the program and 99.9% of beneficiaries could identify the specific climatic shock associated with the payouts they received.

[11] In addition, a complementary insurance policy can be contracted directly by the Ministry of Agriculture for uninsured vulnerable areas. In this case, the areas to be insured are determined in May and state governments can opt to pay their corresponding portion of the premium if they wish to receive indemnity payments directly.
[12] The states can also, after federal authorization, use the funds in alternative ways, such as the re-construction of infrastructure that has been affected as a result of the climatic shock (FAO 2014).

Results on Farmer Behavior and Poverty Reduction

We begin by summarizing several concerns related to contract design expressed by Fuchs and Wolff (2011b), who analyze the Mexican WII contract in detail. First, the current practice of basing the index on cumulative rainfall within a period should be complemented with the variance of rain during the period because yields are affected not only by the total amount of rainfall but also the number of days of rain and the timing of rainfall. Second, thresholds of rain millimeters below which indemnities must be paid should be readjusted over time (thresholds have not been changed since the beginning of the program) so as to avoid inhibiting investment in research for the development of drought resistance seeds. Finally, the authors identify several potential negative spillover effects of the WII contract. These effects include the discouragement of investment in irrigation infrastructure (WII is only available in rainfed areas) and of crop diversification since relatively few crops beyond maize are covered.

However, the same authors (Fuchs and Wolff 2011a) performed an in-depth analysis of the effects of WII on maize productivity, per-capita income and expenditure, and farm-level risk management from 2003 to 2008. Comparing initially treated counties with counties later covered by WII and counties never treated with WII, the authors find a positive effect on maize productivity (6%), which reveals an ex-ante response; that is the index insurance induced farmers to increase input intensity and improved production techniques. The authors also found positive spillover effects: the area planted with maize decreased by 8%, with an expansion in the area devoted to more profitable, commercial crops. This finding mitigates the concern mentioned above about crop overspecialization. Finally, the authors point out that credit constraints were likely reduced by the WII program, a result consistent with the intensification of production and increase in yields.

Index-Based Livestock Insurance (IBLI) in Kenya

IBLI constitutes the first livestock insurance in Africa (Mude 2014).[13] It was designed by the International Livestock Research Institute (ILRI) with Cornell University, Syracuse University and the BASIS Research Consortium as technical partners. IBLI was introduced in northern Kenya in 2010 and then in Ethiopia in 2012. IBLI targets pastoralist households in arid and semi-arid lands. In Kenya, for example, more than three million households depend on livestock as their primary, or in many cases only, asset and livestock generate more than 60% of their income (Chantarat et al. 2013). Severe drought, which is becoming increasingly frequent and unpredictable, is the main cause of livestock mortality in this region and causes significant hardship for pastoralist households. In Kenya, the program started in Marsabit District, where high-quality historical livestock mortality data were available. Based on encouraging findings from impact evaluations in Marsabit, the program was extended to four other districts between 2013 and 2015 (Jensen et al. 2015).

In this section we briefly describe the data used to design the original contract in Kenya, the contract implementation and the main results noted in the impact evaluations.

[13] The policy covers the primary forms of livestock in the region including goats, sheep, cattle and camels.

The Data for IBLI

The IBLI contract was developed using two types of historical data sets: one tracking losses to the primary asset to be insured and one allowing construction of potential indices correlated with these losses. The first data set contains household-level livestock mortality collected via a monthly survey conducted in Marsabit District by the Kenyan government's Arid Land Resource Management Program (ALRMP). These livestock mortality data were collected between 2000 and 2008, allowing the construction of a sample of 112 season-specific observations of average livestock mortality rates across smaller geographic sub-regions within the district (Chantarat et al. 2013). It is important to note, however, that the ALRMP data cannot be used to implement a widely available index insurance contract based on directly measured average mortality rates because the sample is clustered spatially and does not have sufficiently broad coverage across space.

The second data contains the Normalized Difference Vegetation Index (NDVI), an indicator of vegetation density, computed by the National Aeronautical and Space Administration (NASA) using satellite data collected by the National Oceanic and Atmospheric Administration of the United States (NOAA). NDVI is the basis for constructing the contract's index with the logic that vegetative density falls significantly under severe drought as the natural pasture and forage that the pastoralists rely on to feed their livestock declines. The NDVI data are characterized by high spatial resolution (8 km2) and are been generated at a 10–16 day frequency. These data, which are collected by US institutions and freely available, have reliably provided information about Africa's pastureland since 1981 (Mude et al. 2009; Chantarat et al. 2013).

In order to design the IBLI contract, researchers estimated the relationship between various potential NDVI indices and livestock mortality. The index that was eventually selected was the best predictor of severe mortality incidents. IBLI represents an important step forward in the design of index insurance contracts in the developing world. It represents the first index-insurance contract developed on the basis of household-level panel data. These data permitted both careful estimation of the relationship between a range of alternative NDVI-based indices and herd mortality and, importantly, the ex-ante evaluation of the magnitude of basis risk associated with each potential index. This approach, while demanding in terms of data and human capital, is crucial to identify contract structures that minimize basis risk and thus provide maximum value for the insured households (Mude et al. 2009).

The Contract Implementation

In addition to careful contract design, the implementation of IBLI was also carefully thought-through to allow for a rigorous evaluation of the program in order to generate learning about welfare impacts on households and lessons that would help improve the design of the contract itself. The evaluation implied adopting an experimental design of the contract's implementation, including randomly assigned price incentives and extension/educational campaigns. Baseline data on key characteristics of a random sample of households was gathered prior to the initial offering of the IBLI contract in Marsabit. A similar household survey is applied annually to the same households in order to evaluate the impact of this contract on pastoralists' well-being, perform a rigorous analysis of basis risk and understand what motivates pastoralists to purchase the insurance (Jensen et al. 2015). Given the existence of an unconditional cash transfer program in the same district since 2009, the Hunger Safety Net Program (HSNP), the experimental design also allowed for an opportunity cost analysis of the use of public funds in IBLI by comparing its benefit/cost ratio to that of the HSNP (Jensen et al. 2015; Mude et al. 2010).

IBLI's implementation was facilitated by the presence of the HSNP since the cash transfer program had already created the infrastructure for delivering the cash aid in the same remote areas covered by IBLI. Specifically, Equity Bank made a large investment in wireless portable devices (point of sale software or POS system), which allowed to collect IBLI's premium and distribute indemnity payments, thus reducing the need for insurance agents to carry cash and greatly reducing the transactions costs of the insurance (Mude et al. 2009).

Northern Kenya has a bimodal rainfall distribution and two associated growing seasons. IBLI contracts are offered for both seasons (long rain-long dry and short rain-short dry periods). The contracts are sold approximately two months before the beginning of each season, and predicted livestock mortality is announced at the end of each season. The predicted livestock mortality rates are generated by plugging the value of the NDVI index for the season into the livestock mortality model described above. Indemnity payments are made if the predicted livestock mortality exceeds the strike-point of 15%. Payouts to farmers are equal to the difference between the predicted mortality rate and the strike point, times the value of the insured herd. This value, in turn, is the number tropical livestock units (TLU)[14] the household chose to insure times a pre-agreed value per TLU (Chantarat et al. 2013; Mude et al. 2009).

The Results

Researchers have evaluated IBLI based on a number of outcomes including pastoralists' demand for the product, as well as IBLI's impacts on pastoralist's investment behavior and consumption, income and poverty levels. As noted in Jensen et al (2015), there has been strong and growing demand for IBLI since the beginning of the program, with the percentage of eligible households purchasing the insurance increasing from 30% to close to 50% over the first three years. On the other hand, there has also been a growing rate of dis-adoption, that is, households who had purchased the product at least once but that did not buy it any further (from about 20% to close to 40%). Potential reasons for the observed dis-adoption rates include: discouragement due to absence of indemnity payments early on and logistical issues complicating product sales. Other variables found to have a strong influence on demand for IBLI are the product's price, financial liquidity of the household, the level of covariate risk in the region and individuals' predictions about rangeland conditions, as well as the availability of alternative coping strategies (Jensen et al. 2015; Chantarat et al. 2009).

Ex–ante responses to IBLI have included a decrease in herd size,[15] greater investment in veterinary and vaccination services for the livestock, and other changes in production strategies leading to increased milk productivity and improved household income and nutrition (Jensen et al. 2015).

The catastrophic drought in the Horn of Africa in 2011, triggered the first indemnity payouts. A study by Janzen and Carter (2013) identified the ex-post impacts of IBLI on distress

[14] The TLU measure allows for aggregation of different species based on their average metabolic weight. That way, 1 TLU = 1 cattle = 0.7 camels = 10 goats or sheep (Chantarat et al. 2013).
[15] Since households tend to hold assets as precautionary savings, a reduction in herd size and a corresponding increase in consumption can occur as a response to the availability of insurance. This can be expected for households above a critical asset threshold (Janzen et al. 2013).

livestock sales and food consumption. Since their survey was performed at the time of the payouts, they asked households about the way they had been coping with the drought during the three months prior to the survey (Q3), and about the ways they were planning to cope with the drought during the three months after the survey (Q4). Compared to uninsured households, the authors found that insured households were, on average, thirty-six percentage points less likely to resort to livestock sales and twenty-five percentage points less likely to decrease the number of daily meals during Q4.

Interestingly, however, the researchers find a heterogeneity of anticipated reactions to insurance payouts based on an identified critical asset threshold. Economic theory on the accumulation of productive assets predicts that relatively asset rich households will tend to reduce their assets in the event of a shock so as to smooth consumption, while asset poor households will instead hold on to their assets, sacrificing food intake, so as to preserve their limited income-generating capacity. The results of this study support the theory and show positive effects of IBLI on these two types of households. That is, asset rich households were significantly less likely to sell assets during Q4 compared to uninsured ones (this impact was statistically insignificant for insured poor households), hence helping these families to preserve their source of future income.

On the other hand, asset poor households were found less likely to reduce the number of meals thanks to IBLI's payout (this impact was statistically insignificant for insured richer households). This finding implies that IBLI led to a reduction in malnutrition in this food insecure region. Janzen and Carter (2013) also found that IBLI positively impacted asset poor households in Q3, as they had tended to rely less on coping strategies that would destabilize consumption because they expected a payout from IBLI.

Another noted impact of IBLI has been a 33% reduction in food aid needed for northern Kenya (Malone 2014). The positive results obtained from IBLI have encouraged its expansion to other Kenyan districts and also its evolution to be able to rely only on NDVI data so as to be offered in areas with lack of data on livestock mortality (IBLI's website 2015).

The "Shadow" Area - Yield Index Insurance in Ecuador

Ecuador enjoys a privileged situation with respect to the availability of agricultural yield data. Specifically, since 2000, the government of Ecuador has administered the Continuous Area and Agricultural Production Survey, known by its Spanish acronym ESPAC. The ESPAC is a national survey that collects data on area planted and yields and thus can potentially serve as the basis for an area yield index. Unfortunately, while a relatively large quantity of high quality yield data exist, the same is not true with respect to weather data in Ecuador. There are relatively few meteorological stations, including only two automated stations, and the data that do exist are insufficient to design index-based contracts.

In order to explore the viability of index insurance for small-holders in Ecuador, the authors carried out a research project between 2010 and 2014. This project included the design of a hypothetical, or "shadow", area yield index insurance contract and an analysis of the degree to which this shadow contract, had it been available, would have improved the income and consumption smoothing capacity of maize and rice farmers in three separate cantons.[16] The performance of the shadow index contract was then compared to that of the actually existing, conventional insurance contract that was offered in the same areas. The provision of significant premium subsidies for this conventional contract has been the primary government policy to strengthen crop insurance markets since 2010.

The analysis was based on data collected from a panel of 1,000 maize and rice farmers surveyed in 2011 and 2012 in two of the main maize growing cantons (El Empalme and Celica) and one rice growing canton (Daule). All sample farmers were insured under the conventional insurance contract, primarily because they had taken out loans from the formal banking sector, which has increasingly required farmers to hold crop insurance as a condition for credit access.

In the remainder of this section, we describe the ESPAC yield data, the construction of index insurance contract areas and briefly discuss results of the comparison of the two types of contracts (index vs. conventional) in terms of their ability to shield farmer's income from yield risk.

The Historical Data: The ESPAC Yield Survey

The ESPAC is a survey administered annually by Ecuador's National Census and Statistics Bureau (INEC), with the primary objective of generating province-level production and yield estimates for the country's primary crops. The ESPAC uses the 2000 agricultural census as its sample frame. The census divided the country's cultivable land into Primary Sampling Units (PSM), which are contiguous areas of approximately ten square kilometers that are homogeneous in terms of agro-ecological conditions. Each PSU, in turn, was sub-divided into smaller sampling units called Sample Segments (SS) of approximately two square kilometers. In 2002, from the universe of 69,272 SS's throughout the country, INEC randomly selected 2,000 for inclusion in the ESPAC sample. Within these chosen SS's, INEC applies the annual ESPAC survey, which collects information on land use, area planted and production, for all plots within each SS.

Since 2002, INEC has carried out the ESPAC in the same 2,000 SS's each year.[17] Including the data collected from these same SS's in the 2000 census, the ESPAC data set consists of a twelve-year panel of all plots within these 2,000 SS's (i.e., 2000, 2002–2012).[18] These data, collected as part of the government's annual yields survey, permit us to design the "shadow" index contract.

Definition of Contract Areas: Clusters of Sample Segments

With the large government-collected historical yield data in hand, the first step was to design the index insurance contract. Given the large quantity, both cross sectional (i.e., many plots per season) and over time (i.e., from 2000–2012) and high quality of yield data available, we

[16] Canton is the administrative unit below the province in Ecuador.
[17] 2006 was the only exception. In that year, due to a one-time budget expansion, the ESPAC was carried out in 3,610 SMs.
[18] Given that our objective was to construct a shadow index contract for corn and rice, we restricted attention to those SS in which at least one plot was planted in the relevant crop (rice in Daule and corn in El Empalme and Celica) in each year.

constructed a direct area-yield index for our "shadow" contract. A first step in operationalizing the shadow contract is to define the contract area, or the geographic areas in which average, or area, yield is calculated. Once the contract areas are defined, the historic data from the ESPAC yield survey can be used to estimate the probability distribution function of area yield for each contract area.

There are several options for defining the contract areas. At one extreme, we could define the entire canton as a single contract area. Under this option, we would combine the data from all of the SS's within the canton to estimate the average yield in the canton. This option would be attractive if the canton was characterized by a high degree of homogeneity in terms of agro-climatic conditions. Unfortunately, the cantons in our study (and in general in Ecuador) are characterized by a high degree of internal heterogeneity and, as a result, this option would result in a high degree of basis risk.

At the other extreme, we could define one contract area for each SS. While this option would reduce the level of basis risk, it suffers from two potentially serious problems. First, since there are relatively few (between 10–50) plots in each SM, this option would generate an estimate of average yield that is likely to have relatively low statistical precision (i.e., relatively large confidence interval around the estimate). The second concern is more operational since this option would imply defining and executing a different contract for each SS and, as a consequence, would increase the operating costs of the insurance policy. In the case of El Empalme, for example, this option would imply defining thirty-six separate contracts.

The option we chose for this exercise represents a middle ground in terms of spatial aggregation. Specifically, in each canton, we use the statistical technique of cluster analysis to group together similar SS's into a small number of contract areas. We defined clusters that maximize the co-movement between average yields across SM's over the historical period for which we have data from the ESPAC survey: 2000–2012. The result of this statistical procedure was the definition of three contract areas in each of the cantons of El Empalme and Daule and two in Celica. While in some cases the clustered contract areas include SM's that are quite spatially concentrated, in other cases they include SM's that are more distant from each other but that share certain characteristics (for example altitude or bordering a river) that imply a high degree of co-movement in average yields.

In order to evaluate the hypothetical performance of the shadow index insurance contract for our sample, we assigned each plot operated by our 1000 sample maize and rice farmers to the contract area associated with the nearest ESPAC SS.

Results: Comparative Performance of Index versus Conventional Insurance

In order to make a meaningful comparison across the two types of contracts, we chose a strike-point for the shadow index contract such that its price would be the same as that of the actually existing conventional contract. We thus answer the question: Which type of contract offers greater protection f*or a given cost*? Our comparison is based on two alternative measures that influence the degree to which the insurance contract affects farmers' end-of-season income after accounting for premiums paid and indemnities payments received. We are particularly interested in the success of the insurance contracts in maintaining a minimum level of earnings for those farmers who suffered the greatest losses, which is to say those who are in the lower deciles of the yield distribution. The two measures are:

- The net revenues received by the farmers defined as gross revenues minus the premium payment plus any indemnity payment received,
- The fraction of farmers in each decile of the yield distribution who receive an indemnity payment.

For brevity, the focus of these findings is for maize farmers. Ideally, we would like to have a long time series over which to compare the performance of the two contracts. However, we only have information on two years (2011 and 2012). Fortunately, these two years were very different in terms of climate and agricultural production, thus providing a useful window through which to evaluate the quality of the contracts.

Observations from the two years were divided into ten yield deciles. Each decile was then compared for the following three situations: 1) Net revenues per hectare under the "shadow" contract (i.e., gross revenues minus any indemnity payment that would have been received under the index contract scheme minus the hypothetical index premium); 2) Net revenues per hectare if they had no insurance and; 3) Net revenues per hectare under the actual conventional insurance scheme (i.e., farmers' actual net revenues).

It was found that farmers in the lowest yield decile in both 2011 and 2012, the net revenue per hectare was $217 under the existing conventional insurance scheme. Net revenue per hectare would have been would have fallen to $133 under the no-insurance scenario; and would have been $329 under the "shadow" index contract. This result of greater protection offered by the shadow index insurance contract than the conventional contract held over the bottom five deciles of the yield distribution.[19]

This better protection offered by the index contract for the bottom of the yield distribution can be seen clearly in Figure 1. For farmers in the first five deciles—who suffered the greatest losses and needed the largest indemnity payments—the net payment of the index insurance contract (indemnity minus premium) would have been about two times as high as the net payment farmers actually received from the conventional insurance contract. This difference does not necessarily indicate a failure of the conventional insurance, rather it reflects the higher operating costs and the less generous level of coverage. Stated another way, for the same cost, index insurance offers significantly more protection against yield driven income fluctuations.

Turning to the frequency of receiving indemnity payments, the performance of the two contracts is similar. For the bottom yield decile, a bit more than 90% of farmers would have received a payment under index insurance, while slightly below 90% received a payment from the existing conventional insurance. Once again, index insurance dominates the conventional insurance in the first five deciles. In the highest deciles, the realized losses are likely more idiosyncratic and do not reflect events, such as drought, that result in massive losses. In these deciles, we can see the existence of basis risk with the index insurance, although it is clear from Figure 1 that the payments by the conventional insurance are small on average.

In global terms we see that, for the first five deciles the average income of farmers is higher with both the index and conventional insurance contracts than without any insurance.

[19] The maize farmers from the seventh decile on had relatively good yields. As is reasonable, these farmers would have been slightly better off without insurance than with either of the two insurance types.

Nonetheless, we observe that the index insurance contract consistently offers greater protection than the conventional insurance across each of the bottom five deciles. The impact of index insurance is indeed greatest where it is most needed; for the first decile, the index insurance would have increased income by $200 per hectare, while the difference is only $100 for the fourth decile. On the other hand, we observe that the conventional insurance is slightly better than index insurance for the highest deciles (at the most it offers $10 per hectare on average).

Although these results show clear advantages for the area yield-based index insurance developed using the ESPAC data, we need to be cautious given that these results were based on only two years of information, one of which was characterized by a catastrophic drought for maize producers. It is precisely in that year that the protection offered by the index insurance would have been much greater, compared to the more "normal" year in which the losses were likely more the result of idiosyncratic factors, or more routine shocks, which tend to affect few people. In this latter case, conventional insurance provided slightly better protection.[20]

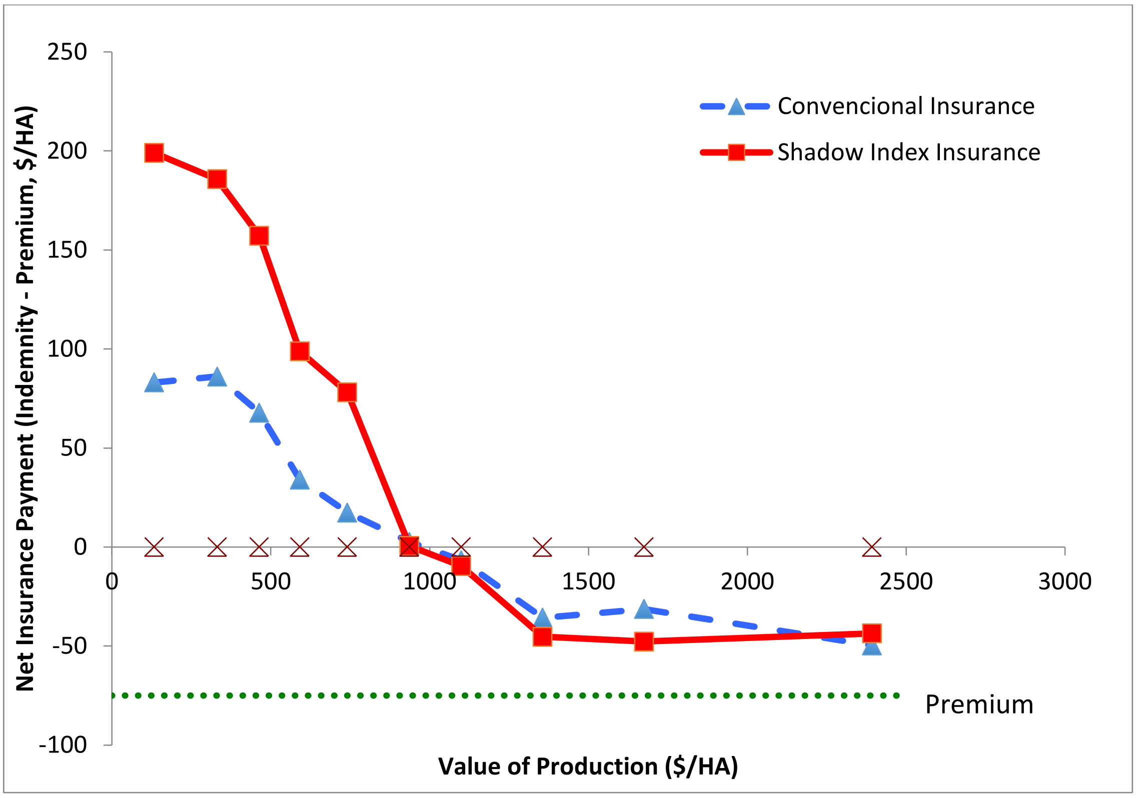


Figure 1. Net payments received by maize producers, 2011 and 2012

Notes. The horizontal axis represents the value of production for the various deciles in dollars per hectare. The 'x' on this axis show the position of the various deciles. The solid red line shows the value of the indemnity minus the cost of the premium for the index insurance. Without any indemnity payment, the line would be at the level of -$75 per hectare (the value of the premium). The blue dashed-line shows the same net impact on income for farmers in the various deciles for the conventional insurance.

[20] About 25% of farmers received payments under the conventional insurance in 2012. These payments were small on average and did not reach the level of the premium in any decile. For both contracts, this is what would be expected in a year without large losses.

Importantly, our study shows that Ecuador's national yield survey (ESPAC) could likely serve as the basis for the development of an area yield index insurance policy that, dollar for dollar, offers better protection for small and medium farmers compared to the existing conventional insurance policy. The area yield index insurance contract offers the potential for better protection not just to individual farmers, but also to protect the portfolios of financial institutions with significant agricultural loan portfolios and, potentially, to local governments in regions with high dependence on agriculture against climate-related disasters. Supporting this type of contract would also imply a more cost-effective use of public funds compared to the current conventional insurance subsidy. The governmental institution in charge of the ESPAC is currently in the process of updating and expanding that survey, a policy that provides a unique opportunity to develop a high quality area yield index insurance market in Ecuador.

Big Data and its potential for Index Insurance

Our discussion suggests that index insurance holds the potential to improve the risk management capacity of small-holder farmers in developing countries and could play an important role in reducing poverty and enhancing rural development. While significant strides have been made, index insurance markets remain thin and, and even where it is available, demand is relatively low. Realizing the full potential of index insurance, through expanded coverage and improved contract design, requires creativity and innovation. Although challenging, increased incorporation of big data in the design, execution and evaluation of index insurance offers an attractive area for innovation and creative thinking.

Big data derives from technological applications, such as cell phone apps, satellite and radar-based imaging and drone-based imaging, that generate unstructured data in high volumes and at high frequency. These data, if structured and analyzed, can be useful for a variety of purposes including marketing, health care, agricultural extension and support, climate predictions, and national security. Structuring and analyzing these data is, however, not an easy task. It requires powerful analytical tools that allow rapid, high-frequency analysis and high quality human resources with sufficient statistical knowledge and the ability to work with these tools and interpret the results (Sonka 2014; Manyika et al. 2011; da Silva 2016).

A primary challenge to generating and using big data in developing countries is insufficient access to technology, particularly the requisite computing power, internet bandwidth and sophisticated software (da Silva 2016). Another major challenge is the lack of analysts with the skills described above. While fully overcoming these limitations will require time and long term investment in human capital, a number of strategies, including developing key public-private partnerships, could be implemented in the short term in order to speed developing countries' capacity to benefit from big data.

Initiatives like the recent partnership between Google and FAO, aimed at facilitating developing countries' access to satellite data in order to improve their capacity to plan and monitor the use of their natural resources, represents one example of this type of partnership (FAO 2016). Through this partnership, FAO's offices in member countries can request training of their staff and technical experts to use Google technology to access and analyze satellite data for identified needs such as monitoring deforestation rates, carbon sequestration, and agricultural yields. This type of collaboration can help public and private institutions in developing countries access and effectively use the copious amounts of meteorological and agricultural data available through big data to both improve the quality of existing index insurance contracts (i.e., reducing basis risk) and expand coverage of index insurance to currently unserved areas. The Radar-based remote sensing Information and Insurance for Crops in Emerging economies (RIICE) project is one example of a collaboration that is putting these ideas into practice. Five partners[21] have joined together to make use of radar-based remote sensing technology (or Synthetic Aperture Radar–SAR) to provide information on rice growth in Asian countries to enhance food security and strengthen insurance markets (RIICE 2016; Holecz et al. 2013). RIICE takes advantage of data collected by radar sensors in satellites of the European Space Agency and other providers. Because these sensors can detect vegetation growth without the need of direct observation (i.e., they are not restricted by cloud cover), this collaboration permits the use of remote sensing data in the design of index contracts in areas, such as the highland and jungle regions of Ecuador, where dense cloud cover throughout the year has previously ruled out the development of satellite based insurance contracts. Local public sectors play a key role in the partnership, which has been implemented in parts of six Asian countries since 2012, by participating in product development and gathering terrestrial data for validation or "ground truthing" of the satellite data (Holecz et al. 2013). Based on these validation exercises, the accuracy of the estimates of planted area and rice yields generated by the RIICE project is significantly higher than the conventional estimates generated by national statistical offices.

Another important advantage of the RIICE estimates is the speed of generating actionable data. For example, RICIE estimates of crop yields or crop losses are available within several days, thus allowing governments or insurance companies to respond to catastrophes in a much more timely manner (ASEAN SAS 2016). This project is now in its second stage (2015–2018), which includes the piloting of national crop insurance programs (Ibid).

In a related partnership, the Global Index Insurance Facility (GIIF) of the World Bank and AXA Corporate Solutions (AXA CS) have joined forces to promote the use of big data as a means of extending weather index insurance to regions that were previously uninsurable because of low quality or lack of weather data. Researchers in this partnership are facilitating government and private insurance sector access to and management of satellite data in order to generate higher quality (lower basis risk) weather index insurance contracts compared to contracts based on more limited data from meteorological stations (AXA 2015).

A range of additional options in the form of new technological platforms and devices exist to promote the generation and use of big data in index insurance. One particularly promising example is the use of drones to generate high quality yield data at a spatial resolution that is sufficiently high to write area yield index contracts. For example, in the case of Ecuador, implementing an area-yield index insurance contract using a survey like ESPAC could be complemented by the use of drones, which can provide high quality images that permit precise monitoring of crop development throughout the agricultural season. From the images, data can be gathered and processed to support ESPAC's findings and to generate feedback for improving contract design, for example by providing an "audit" in cases where a high percentage of farmers within a contract area suffer a loss but the insurance contract, according to the value of the index, would not normally pay out. In the long run, the use of drones could potentially serve as the primary means of generating yield data for a range of purposes, including the development and implementation of index insurance contracts.

[21] The five partners are the German Development Cooperation, the Swiss Agency for Development and Cooperation, the International Rice Research Institute, the Allianz Re insurance company, and Sarmap, a software provider.

While big data presents multiple possibilities for designing better insurance contracts and offering contracts where they were previously infeasible, it can also facilitate the implementation, monitoring and evaluation of index insurance programs. In the case of Mexico described above, for instance, the development of mobile money or "electronic wallets" with which farmers can carry out transactions through their cellphones, can be used to transfer indemnity payments to farmers in remote areas (and collect insurance premiums from farmers if it were the case), thereby dramatically reducing the transaction costs associated with implementing insurance. The big data generated by these transactions could also be used to monitor how the indemnity payments received are spent. E-wallets have been successfully introduced in a number of low-income countries including Nigeria—for the transfer of government subsidies (Akinboro 2014)—and Kenya—for microfinance-loan repayments (The Economist 2013). By allowing and recording millions of transactions in the rural sector, big data can be analyzed to gain a better understanding of how access to insurance affects key farmer behavior such as the purchase of improved seeds and fertilizer as well as household's consumption patterns. Cellphones can also be used to notify farmers about the availability of insurance programs and educate farmers about the costs and benefits of insurance so that they make informed demand decisions. The GPS location of cellphone holders can also help governments and insurance providers monitor the number of policy holders and program beneficiaries in affected locations and thus potentially monitor disaster occurrence and relief by region.

Kenya, the home of the IBLI program discussed above, is a global leader in mobile money (The Economist 2013). In Kenya, however, a number of challenges currently limit the spread and use of e-wallets, including limited access to cellphones and energy for charging the phones, as well as limited availability of mobile money network agents (Hanson 2014). These limitations are being resolved with time (Ibid), suggesting a tremendous potential for cellphones and their associated rapidly expanding services, such as mobile money, to promote the deepening of financial services, such as credit, savings and insurance via big data. In the specific case of IBLI, monitoring the use of indemnity payments through e-wallets may prove to be a valuable part of the impact evaluation strategy in the near future.

Also important to note is, as observed by da Silva (2016), the potentially important role of cooperatives for data access and data collection, especially in the context of small farmers. The organization of small farmers may both facilitate technology adoption as well as encourage the exchange of information both among farmers and between them and national research institutes or public/private service providers.

Conclusions

This paper has emphasized the key role of publicly available weather and agricultural yield data in the design of index-based insurance schemes for small-holder farmers in developing countries; precisely the segment of the rural population that is almost universally uninsured and for whom covariate risk can create and perpetuate poverty traps.

The Mexican and Kenyan cases summarized here show how index insurance has helped poor farmers by encouraging more efficient productive behavior and by reducing the use of costly ex-post risk coping strategies such as asset depletion or consumption destabilization. These benefits have led to higher per-capita income and to a reduction in malnutrition.

The Ecuadorian case also illustrated the potential of index insurance to smooth farmers' income in the face of covariate shocks such as droughts. Indeed the shadow index contract developed by the authors performed more favorably (albeit under hypothetical circumstances) than conventional named-peril insurance contracts available to farmers (in the real world). These results suggest that the development of area-yield based index insurance in Ecuador would likely represent a better use of public funds than the continued subsidization of conventional insurance.

The Mexican and Kenyan cases also highlight the savings afforded to the public budget by the implemented index insurance programs; in the case of Mexico by reducing direct support and disaster relief payments in the wake of catastrophic weather events and in the case of Kenya by reduced alliance on international funds for food aid.

The portrayed cases also suggest that index insurance is both a dynamic field and a field that has significant scope for improvement moving forward. Contract design can be improved and basis risk reduced by using new data sources that permit alternative indices or by more effective use of existing data (following, for example, the observations of Fuchs and Wolff (2011b) for Mexico's catastrophic weather index insurance). Creative use of satellite imagery (as in the case of the IBLI contract in Kenya) has recently allowed governments and the private sector to extend insurance coverage to previously uninsurable areas. The big data revolution (i.e., the availability of higher quality data and more developed computing power and methods for analysis) is likely to significantly increase the dynamism of the index insurance sector by further expanding the types of data that can be used to design indices and extending geographic coverage. The challenges implicit in the successful utilization of big data in developing countries, however, suggest that the development of public-private partnerships will be crucial in order to take advantage of the potential offered by big data.

All indices require reliable and long series of data. However, the absence of data need not prevent the development and implementation of index insurance initiatives. Strategies such as public-private partnerships for insurance or reinsurance, and hybrid combinations of multiple indices as well as indices combined with on-farm yield audits should be considered. Instead of a vicious cycle that discourages the development of index insurance, innovative efforts such as those discussed here to overcome data limitations can create a virtuous cycle in which the productive use of information, (i.e. the development of index insurance schemes), encourages further public (and also private) investment to improve the timing, quantity and quality of data collection. This virtuous cycle can lead to the creation and expansion of a sustainable insurance market that can effectively add value to farm businesses, lending institutions and other service providers along the agricultural supply chain.

References

AGROASEMEX. 2006. La Experiencia Mexicana en el Desarrollo y Operación de Seguros Paramétricos Orientados a la Agricultura. [The Mexican Experience in the Development and Operation of Agricultural Parametric Insurance] AGROASEMEX S.A, Mexico.

AGROASEMEX. Productos y Servicios. http://www.agroasemex.gob.mx/ProductosyServicios/ Seguros.aspx#horizontalTab3 [Products and Services] [accessed October 12, 2015].

Akinboro, B. 2014. Bringing Mobile Wallets to Nigerian Farmers. Blog at CGAP–Advancing Financial Inclusion to Improve the Lives of the Poor. http://www.cgap.org/blog/bringing-mobile-wallets-nigerian-farmers.

ASEAN Sustainable Agrifood Systems. 2016. http://www.asean-agrifood.org/projects/riice/ [accessed March 4, 2016].

AXA. 2015. Parametric Insurance: A Fitting Solution for the Weather-Sensitive - A Look at New Weather Risk Management Solutions to Protect People, Companies and Economies. AXA Corporate Solutions. https://www.axa-corporatesolutions.com/IMG/pdf/15_254_AXA_brochure_170x240_EN_v2.pdf.

AXA. 2015. Announcement of a Partnership between AXA and the World Bank Group: Insurance and Big Data, Innovating to Face Climate Change. Press Release: 17-Febuary 2015. https://group.axa.com/en/newsroom/news/news-insurance-and-big-data.

Barnett, B., C. Barrett, and J. Skees. 2008. Poverty Traps and Index-Based Risk Transfer Products. *World Development* 36 (10): 1766–1785.

Binswanger-Mkhize, H. 2012. Is There Too Much Hype about Index-based Agricultural Insurance? *Journal of Development Studies* 48 (2): 187–200.

Boucher, S., Carter, M. and C. Guirkinger. 2008. Risk Rationing and Wealth Effects in Credit Markets: Implications for Agricultural Development. *American Journal of Agricultural Economics* 90 (2): 409–423.

Breustedt, G., Bokusheva, R. and O. Heidelbach. 2008. Evaluating the Potential of Index Insurance Schemes to Reduce Crop. *Journal of Agricultural Economics* 59 (2): 312–328.

Carter, M. 2012. Designed for development impact: Next generation approaches to index insurance for small farmers. In *Microinsurance Compendium*, Vol II, edited by C. Churchill and M. Matul . Geneva: International Labor Organization.

Carter, M. 2013. Sharing the Risk and the Uncertainty: Public-Private Reinsurance Partnerships for Viable Agricultural Insurance Markets. Policy Brief (78). Fondation pour les Études et Recherches sur Le Développement Inernational.

Carter, M., de Janvry, A., Sadoulet, E. and A. Sarris. 2014. Index-based Weather Insurance for Developing Countries: A Review of Evidence and a Set of Propositions for Up-Scaling. Working Paper (111). Fondation pour les Études et Recherches sur Le Développement Inernational.

Carter, M., Galarza, F. and S. Boucher. 2007. Underwriting Area-Based Yield Insurance to Crowd-in Credit Supply and Demand. Agriculture and Resource Economics Department Working Paper (07–003), University of California at Davis.

Chantarat, S., Mude, A. and C. Barrett. 2009. Willingness to Pay for Index Based Livestock Insurance: Results from a Field Experiment in Northern Kenya. Cornell University.

Chantarat, S., Mude, A., Barrett, C. and M. Carter. 2013. Designing Index Based Livestock Insurance for Managing Asset Risk in Northern Kenya. *Journal of Risk and Insurance* 80 (1): 205–237.

da Silva, J. G. 2016. Big Data, Small Farms: The Role of Data and Statistics in Unleashing the Potential of Africa's Smallholders and Family Farmers. In Special Issue: *African Farmers in the Digital Age: How Digital Solutions can Enable Rural Development. Foreign Affairs*. https://www.foreignaffairs.com/sponsored/big-data-small-farms.

Elabed, G. and M. Carter. 2014. Ex-ante Impacts of Agricultural Insurance: Evidence from a Field Experiment in Mali. Working paper, University of California at Davis.

FAO. 2014. La Gestión de Riesgos Climáticos Catastróficos para el Sector Agropecuario en México: Caso del Componente para la Atención a Desastres Naturales para el Sector Agropecuario. [Catastrophic Weather Risk Management for the Agricultural Sector in Mexico: Case of the Natural Disaster Attention for the Agricultural Sector Program Component] Secretaría de Agricultura, Ganadería, Desarrollo Rural, Pesca y Alimentación, FAO.

FAO. 2016. News Article: Google and FAO partner to make remote sensing data more efficient and accessible http://www.fao.org/news/story/en/item/350761/icode/.

Fuchs, A. and H. Wolff. 2011a. Drought and Retribution: Evidence from a large scale Rainfall-Indexed Insurance Program in Mexico. University of California at Berkeley, University of Washington.

Fuchs, A. and H. Wolff. 2011b. Concept and Unintended Consequences of Weather Index Insurance: The Case of Mexico. *American Journal of Agricultural Economics* 93 (2): 505–511.

Hanson, S. 2014. Can Mobile Money Extend Financial Services to Smallholder Farmers. Blog at CGAP – Advancing Financial Inclusion to Improve the Lives of the Poor. http://www.cgap.org/blog/can-mobile-money-extend-financial-services-smallholder-farmers

Hazell, P., Anderson, J., Balzer, N., Hastrup-Clemmensen, A. Hess, U. and, F. Rispoli. 2010. The Potential for Scale and Sustainability in Weather Index Insurance for Agriculture and Rural Livelihoods. International Fund for Agricultural Development and World Food Programme, Rome.

Holecz, F., M. Barbieri, F. Collivignarelli, L. Gatti, A. Nelson, T. Setiyono, M. Boschtti, G. Manfron, P. Brivio, E. Quilang, M. Obico, V. Minh, D. Kieu, Q. Huu, T. Veasna, A. Intrman, P. Wahyunto, and S. Pazhanivelan. 2013. An Operational Remote Sensing Based Service for Rice Production Estimation at National Scale. ESA Living Planet Symposium. http://www.rcmrd.org/docs/RIICE.pdf.

IBLI. About us. http://ibli.ilri.org/index/ [accessed October 28, 2015].

Janzen, S. and M. Carter. 2013. After the Drought:The Impact of Microinsurance on Consumption Smoothing and Asset Protection. Working Paper (19702). National Bureau of Economic Research.

Jensen, N., C. Barrett, and A. Mude. 2015. The Favourable Impacts of Index-Based Livestock Insurance: Evaluation Results from Ethiopia and Kenya. Research Brief (52). ILRI.

Karlan, D., R. Osei, I. Osei-Akoto, and C. Udry. 2012. Agricultural Decisions after Relaxing Credit and Risk Constraints. Working Paper (18463), National Bureau of Economic Research.

Malone, C. 2014. Insurance Protects Kenya Farmers from Drought. *Al Jazeera*. May 26. http://www.aljazeera.com/video/africa/2014/03/insurance-protects-kenya-farmers-from-drought-201432674122605362.html

Manyika, J., M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburg, and A. Byers. 2011. Big Data: The Next Frontier for Innovation, Competition, and Productivity. McKinsey Global Institute.

Miranda, M. 1991. Area-Yield Crop Insurance Reconsidered. *American Journal of Agricultural Economics* 73(2): 233–242.

Mude, A. The Origins of IBLI and Preliminary Field Testing. 2014. A Film from the International Livestock Research Institute, October 9, 2014. http://iblicasestudy.review/.

Mude, A., C. Barrett, M. Carter, S. Chantarat, M. Ikegami, and J. McPeak. 2009. Index Based Livestock Insurance for Northern Kenya's Arid and Semi-Arid Lands: The Marsabit Pilot. Department of Applied Economics and Management. Working Paper, Cornell University.

Mude, A., S. Chantarat, C. Barrett, M. Carter, and M. Ikegami. 2010. Insuring Against Drought-Related Livestock Mortality: Piloting Index Based Livestock Insurance in Northern Kenya. Economics Faculty Scholarship Paper (75). Syracuse University.

RIICE. 2016. http://www.riice.org/ [accessed March 3, 2016].

Skees, J., A. Goes, C. Sullivan, R. Carpenter, M. Miranda, and B. Barnett. 2006. Index Insurance for Weather Risk in Lower Income Countries. Report prepared for USAID.

Sonka, S. 2014. Big Data and the Ag Sector: More than Lots of Numbers. *International Food and Agribusiness Management Review* 17 (1): 1–20.

The Economist. 2013. Why does Kenya Lead the World in Mobile Money? The Economist Explains. http://www.economist.com/blogs/economist-explains/2013/05/economist-explains-18.

Universidad Autónoma Chapingo. 2010. Programa de Atención a Contingencias Climatológicas (PACC), Evaluación Externa de Resultados. Mexico. [Program for Climatologic Contingencies (PACC), External Results Evaluation].

International Food and Agribusiness Management Review
Special Issue - Volume 19 Issue A, 2016

Organized Data and Information for Efficacious Agriculture Using PRIDE™ Model

Minal Sawant[ⓘa], Rajesh Urkude[b], and Sandip Jawale[c]

[a] *Domain Consultant, Digital Farming, Tata Research Design and Development Centre (TRDDC), Hadapsar Industrial Estate, Pune – 411013, Maharashtra, India*

[b] *Head – Digital Farming Planning and Operations Tata Consultancy Services Limited, Yantra Park - (STPI), 2nd Pokharan Road, Opp HRD Voltas Center, Subash Nagar, Thane - 400601, Maharashtra, India*

[c] *Digital Farming Project Manager, Tata Consultancy Services Limited*

Abstract

Farmers in India are faced with a myriad of issues from access to agricultural inputs, scientific agriculture practices, and market intelligence over climate and weather calamities which often make farming an unprofitable business.

A combined approach of group farming and effective farm management with the help of efficient data collection, processing and analysis is a widely accepted solution to these issues. Progressive Rural Integrated Digital Enterprise (PRIDE™) is an innovative business model which enables rural India to tackle these challenges and prosper collectively. The technology enables efficient collection of data from farmers' fields, agricultural universities, and other private and public stakeholders which is processed and disseminated to farmers on their mobile phones.

Keywords: Information and Communication Technology (ICT) for agriculture, PRIDE, mKRISHI - mobile agriculture, agricultural data management, FPO, group farming

ⓘCorresponding author: Tel: + 91 8976498405
Email: M. Sawant: minal.sawant@tcs.com / minal1689@gmail.com

Introduction

According to World Bank (2016), "Agricultural development is one of the most powerful tools to end extreme poverty, boost shared prosperity and feed 9 billion people by 2050. Growth in the agriculture sector is about two to four times more effective in raising incomes among the poorest compared to other sectors" and can be greatly enhanced through recent cost effective technology developments.

Agriculture in India and Problem Statement

Agriculture, with allied sectors, is unquestionably the oldest and largest livelihood provider in India. It contributed approximately 13.9% of India's GDP (Gross Domestic Product) during 2013–2014; providing livelihood to nearly 600 million Indians (MoA GoI 2015 and DAC 2014). Various studies show that population growth is creating pressure on agriculture to meet the growing demand for food, consequently, leading to rising food prices and poverty levels (ICT in Agriculture 2012; World Bank 2011). While the population of India continues to rise, agricultural productivity is not keeping pace. Farmers face a plethora of problems which are restricting the growth of rural agrarian economies and decreasing the share of agriculture in India's GDP continuously from 1950 to 2014 (Planning Commission GoI 2015).

There is a need to focus on increasing agriculture production in sustainable ways to fulfill the growing needs of the population. Table 1 shows a sampling of vegetable productivity compared with the highest productivity worldwide. Except for a few vegetables, productivity in India is lower than the global average; and in all cases, it is lower than the maximum productivity that can be achieved. This low productivity is due to the lack of access to scientific agricultural advisories, timely availability of inputs, credit, weather information and farm labor (as agricultural labors are migrating to cities for better employment opportunities); and lack of agro-climatic focus in crop selection and management issues (Figure 1). Many small and marginal farmers are attempting to leave farming as the costs of production are higher than the net returns making it unprofitable.

Table 1. Comparative analysis of vegetable productivity in India and worldwide (2012–2013).

Vegetable	Highest Productivity	Productivity in India	Average World Productivity
Tomato	Spain (74 t/ha)	20.7 t/ha	32.8 t/ha
Cabbage	Japan (66 t/ha)	22.9 t/ha	27.7 t/ha
Cauliflower / broccoli	Pakistan (24.8 t/ha)	19.6 t/ha	6.9 t/ha
Okra	Saudi Arabia (13.3 t/ha)	12.1 t/ha	6.9 t/ha
Onion	Turkey (30.3 t/ha)	16.0 t/ha	19.1 t/ha
Potato	USA (44.3 t/ha)	22.8 t/ha	17.7 t/ha
Brinjal	Egypt (49.2 t/ha)	18.6 t/ha	25.0 t/ha

Source. National Horticultural Board, 2013.

Figure 1. Spectrum of challenges faced by farmers in India

Motivation

Globally, researchers are seeking solutions to problems faced by farmers and how to make farming a more profitable venture. Studies conducted by Ghoge (2013); and Gupta and Parida (2013) found that utilizing group approaches to addressing the organization and management of farm activities is an effective problem solving measure. Centralizing agricultural data and information is a key to efficient data management and processing. It helps decision makers make appropriate choices, plan agricultural activities and take preventive and curative measures as needed. ICT tools can provide an important mechanism in achieving the aim of effective data management, handling, processing and dissemination resulting in increased productivity, minimized risks, increased returns from agriculture and ultimately better living in rural areas.

ICT in Agriculture

Information and communication have always mattered in agriculture. Throughout history people have sought information from each another in order to improve efficiency. New advancements in ICT, organizing and processing large amount of data; as well as addressing and managing large farmer groups is becoming more proficient and effective.

ICT tools allow the exchange or collection of data through interaction or transmission. ICT is an umbrella term that includes anything ranging from radio to satellite imagery to mobile phones or electronic money transfers. Advances in affordability, accessibility and adaptability have resulted in large scale use among rural homesteads relying on agriculture. Many of the questions asked by farmers can now be answered faster, with greater ease, and increased accuracy. These types of ICT-enabled services are useful to improving the capacity and livelihoods of poor smallholders and are growing quickly with the booming mobile, wireless and internet industries (World Bank 2011). There are a number of initiatives on the market using ICT-based innovations in agriculture. An analysis of some important advancements are presented in Table 2.

Table 2. Analysis of important ICT product and services interventions in agriculture

ICT Product / Service	Description	Communication Mode	Limitation
Reuters Market Light (RML)	▪ Provides daily information on commodities prices, weather, and advisory services ▪ Services are available in English and regional languages ▪ Network agnostic	SMS and Mobile app messages	▪ Voice messages are not available ▪ Generalized information ▪ Only market intelligence is available, lack of focus on establishing market linkages
IKSL: Indian Farmer's Fertilizer Cooperative (IFFCO), Kisan Sanchar Limited	▪ Joint venture between the telecom network operator Airtel and IFFCO ▪ Information on crops, diseases, weather, and market prices ▪ Dedicated agricultural help line ▪ Information on the availability of products such as fertilizer	Voice based service on mobile	▪ SMS and mobile apps are not available ▪ Generalized information ▪ Lack of focus on establishing market linkages
Farmer's Friend Google product in Uganda	▪ By Grameen Foundation's AppLab ▪ Weather forecasts and agricultural advice ▪ Google trading service for agricultural commodities, and other products ▪ On-demand service (pay at that time, not prepaid) ▪ Generates employment among farmers by hiring some of them for data collection	Mobile App	▪ Works with only Mobile Network Operator MTN Uganda ▪ Voice messages are not available ▪ Generalized information
Digital Green, India	▪ Disseminates targeted agricultural information to small-scale and marginal farmers through digital video ▪ Works with existing, people-based extension systems to amplify their effectiveness	Video	▪ Focus is on dissemination of best practices only ▪ Only static information
e-Choupal	▪ Price information, options for selling the produce, buy inputs at kiosk, advice on farming practices related to input use ▪ Wide spread network	Kiosk and Mobile phone	▪ Generalized information ▪ Crop specific advisory is not available
M-PESA, Kenya	▪ Pilot was focused on microloans and repayments ▪ Person-to-person business model in which customers can buy e-money from agents ▪ Perform financial transactions	Mobile Phones	▪ Only focus on financial transactions ▪ No emphasis on agriculture information and advisory
Esoko	▪ Market information service providing price information and a virtual marketplace for buyers and sellers	Mobile phones (SMS) and Internet	▪ Focus is on market only

Source. IKSL (2016); RML (2016); The Guardian (2013); e-Agriculture (2012); and World Bank (2011).

However, there are limitations with current products on the market. These include:

1. Fragmented approaches to solving challenges in agriculture

The challenges in agriculture are more or less linked. For example if someone is providing scientific agro-advisory to farmers for increasing production, it is equally important for farmers to avail the required agricultural inputs at the prescribed time—at a reasonable rate. Many models provide market intelligence information but there also needs to be a mechanism linking farmers with respective buyers or markets. These services cannot be fully utilized if farmers are not able to act upon it. To maximize the returns from agriculture, problems need to be solved in an integrated manner. A study conducted by Kumar (2011) shows (Table 3) that farmers in India are willing to pay for agricultural services from agro-advisories for market intelligence and prioritize the rankings for services.

Table 3. Priority of farm information

Type of Farm Information	WTP	P &R	WTP	P &R	WTP	P &R
	Uganda		Indonesia		India	
Package of Practices	No	5	Yes	2	Yes	1
Package of Practices (leading to certification)	Yes	3	Yes	3	NA	NA
Pest Information, Alerts & Remedy	Yes	3	Yes	1	Yes	3
Weather Forecasts & Alerts	No	6	Yes	6	Yes	4
Market/Price Information for Commodities	Yes	1	No	7	Yes	5
Access to experts in real time (farm advisory)	Yes	4	Yes	4	Yes	2
Information on Farm credit & subsidies	No	2	Yes	5	No	6

Note. WTP: Willing to Pay**;** P & R: Priority and Rank
Source. Kumar (2011).

2. Lack of Integration among Technologies (Mobile: voice, messages, GPRS; web, etc.)

A mixed approach utilizing technologies for dissemination of vital data and information is necessary to reach a maximum number of users. Kumar (2011) shows (Table 4) that the mobile use among farmers in various countries differs substantially. In India, 90% of mobile users were able to make calls, however only 12–15% can send and receive SMS; and only 2% can access the internet. The percentages vary in Uganda and Indonesia. Although the percentages may have increased in different categories of mobile uses, some users find one feature more usable than another. Further, the cost of GRPS (mobile apps) is lower than the cost of voice messages and SMS. However, GPRS (mobile apps) adaptability in farmers is not so high. A mix of technologies is necessary in order to reach a wider audience and be cost-effective. Moreover, the messaging needs to be in local/regional languages for user understandability and friendliness (Table 5). For other stakeholders (field assistants—experts—FPO management), mobile and web solutions are needed to efficiently collect, process, analyze and effectively convey this useful data and information to farmers.

Table 4. Uses of mobile phones

Activity	Percentage of users who use mobile phones for the purpose in a week		
	Uganda	Indonesia	India
Make phone calls to other mobile phones or fixed lines	72%	82%	90%
Send/Receive SMS from another user	68%	85%	15%
Conduct financial transactions	23%	12%	0%
Listen to music/radio	37%	48%	12%
Click pictures and send to another user	12%	34%	10%
Receive SMS information from operator/third party sources	18%	28%	12%
Access mobile internet (GPRS/CDMA) and 3G	4%	18%	2%

Source. Kumar (2011).

Table 5. English language capacity

Self-reported ability to read English text	Percentage of users using mobile phones for sending and receiving SMS		
	Uganda	Indonesia	India
Not at all	6%	14%	8%
Not easily	14%	20%	12%
Easily	76%	24%	12%
Prefer local language	4%	42%	68%

Source. Kumar (2011).

3. Lack of personalized and real time information

Real time information sharing between farmers and researchers enables service providers to supply real time and personalized services based on a wide range of factors such as: location, crop, management practices, mechanization level, irrigation type, farm size and soil type (Vodafone Group 2015). This allows farmers to make informed choices and take swift agriculture actions when necessary. For example in cases of cyclonic or unexpected precipitation, the real time information helps farmers to make decisions such as whether to prepone harvest by a few days, and thereby avoiding huge losses. This could protect farmers from their season-long efforts and hard work. Accenture Digital (2015) has proven that for developing countries and smallholder farmers, personalized and real time information solutions can enable them to boost field productivity by providing fertilizer, pesticide, and seed recommendations personalized for each farmer's field.

Thus, farmers need an integrated solution involving a variety of technologies. After extensive research, Tata Consultancy Services (TCS) developed an innovative ICT-based platform (PRIDE™) to support the farming community. The model is available in the local language of the user (in this study *Marathi*); and uses various modes of dissemination viz. mobile (voice messaging, SMS and GPRS enabled apps) and web modules. It also accounts for farmer plot and crop specific information, crop history, weather (past, current and forecast), market intelligence, etc. by providing personalized and real time advisory.

Objectives

The objective of this study is to examine the role that the ICT-based PRIDE model has in improving agricultural productivity, reducing production costs, minimizing risks, and ultimately increasing agricultural returns for farmers in one of the project implementation areas.

Specific objectives include:

- To digitize and process farmer, plot, crop and allied data
- To provide personalized agricultural advisories; broadcasts and alerts (weather and market intelligence) to member farmers
- To provide access to agricultural inputs, credit, and access to markets to member farmers
- To increase productivity, optimize cost on inputs, minimize the risks and thereby increase returns from agriculture

Approach and Methodology

Study Area

The selected study area (Figure 2) of thirty-seven villages falls within Dindori tehsil of the Nashik district of Maharashtra state. The economy of Dindori tehsil is primarily agrarian with 64.68% of the total population depending upon agriculture for its primary livelihood (Government of Maharashtra 2014). The total net sown geographical area of Dindori tehsil's is 1342.19 sq. km. or 54.6%. Major crops cultivated in Dindori include: tomatoes, capsicum, grapes, wheat, and onions. The cropping pattern of these crops is depicted in Table 6.

The study area is characterized by semi-arid tropical conditions with an average annual rainfall of 697.6 mm. occurring during the southwest monsoon season (June to September) (Pagar 2012). The mean temperatures range from 23°C to 40°C. It is drained by Godavari River and its tributary Kadwa (Shodhganga 2013). The net irrigated area is around 6% and availability of water is a major problem during the hot season. Soil in this area is derived from Deccan basalt, with a pH of 7.4 to 8.2, containing less clay and silt but rich in organic matter (Shodhganga 2013).

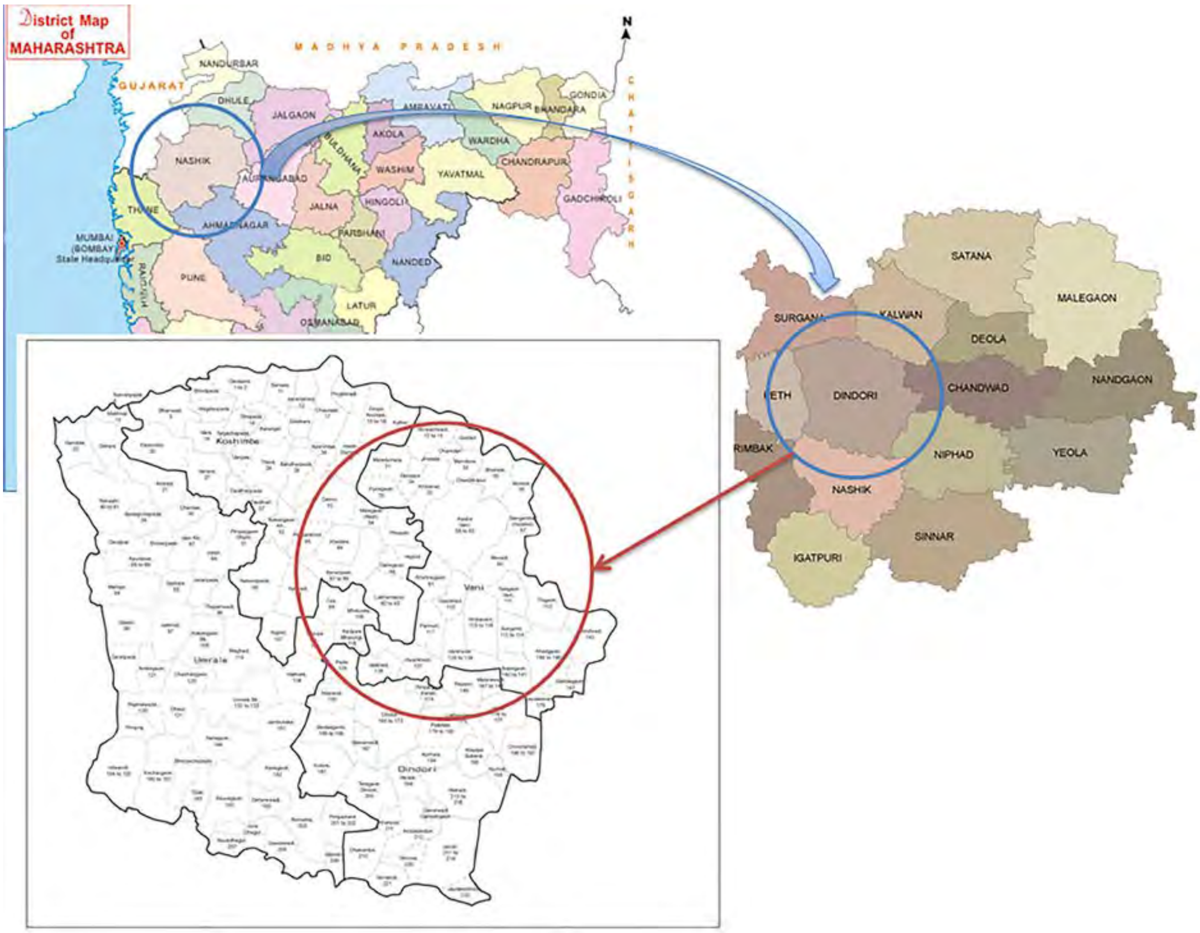


Figure 2. Study area

Table 6. Cropping pattern of major crops in the study area

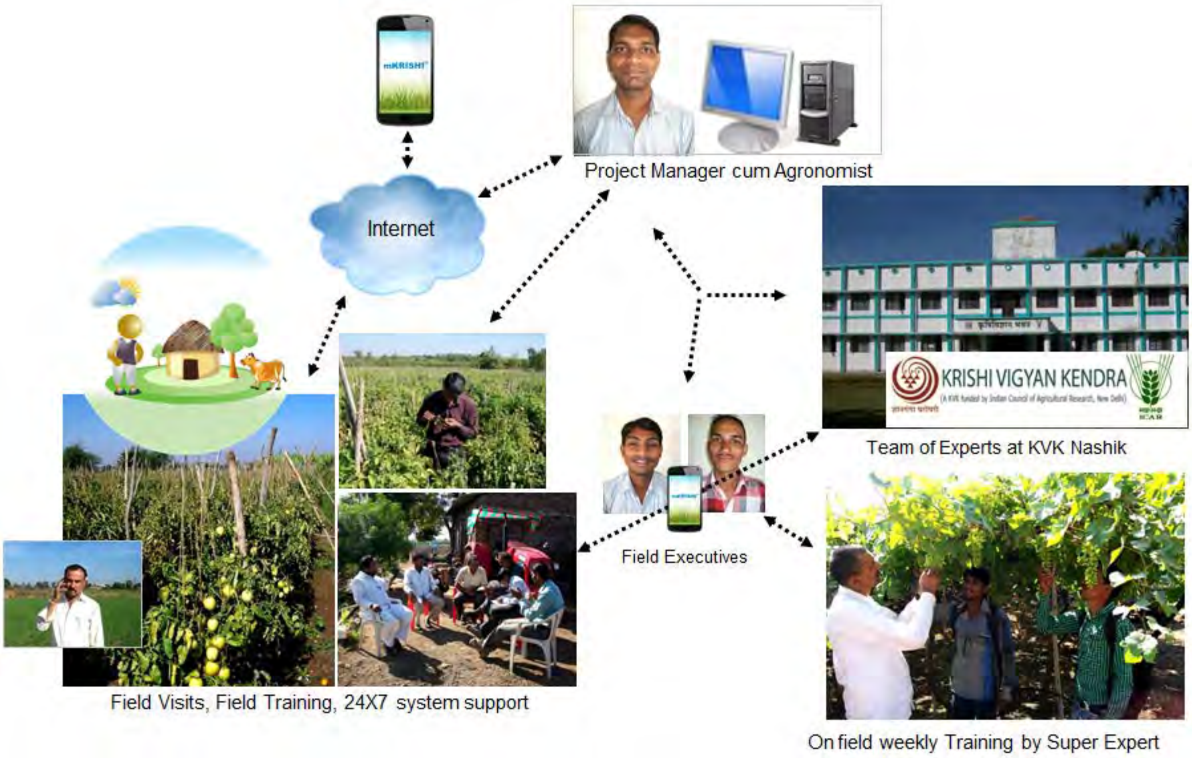
Crop	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Tomato												
Onion (Kharif)												
Onion (Rabbi)												
Onion (Summer)												
Capsicum, Picador Chili												
Wheat												
Grape												
Sowing/Transplant/Pruning			Growth Period						Harvesting			

Tata Consultancy Services' PRIDE™ Platform

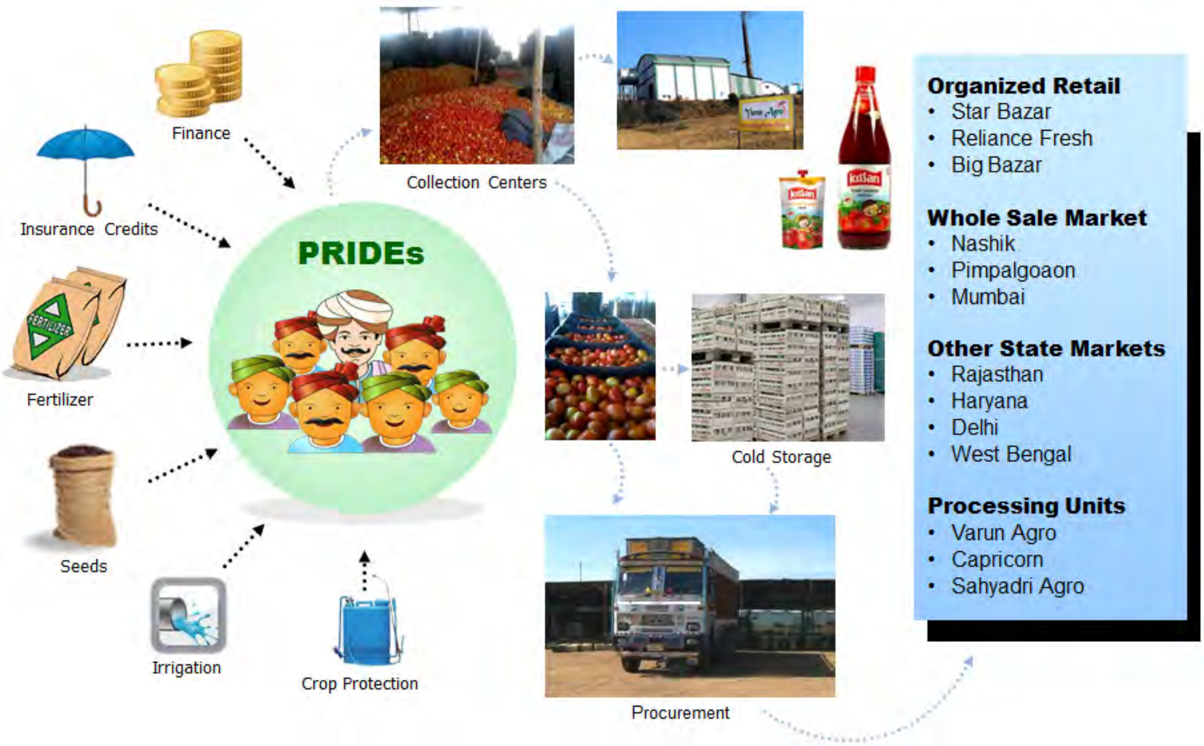
The Progressive Rural Integrated Digital Enterprise, PRIDE™, (more in Appendix 1) is an innovative business model enabling rural Indian farmers to improve farm efficiency through technological interventions (mKRISHI) and collective group management.

Farmer organizations or cooperatives convert to PRIDE through two phases. The first phase is through *Training and Capacity Building* in which farmers, farmer groups, and cooperatives are trained in the mKRISHI modules followed by the digitalization of vital information into the system through the collective efforts of the group. The second phase is the *End-to-End*

Integration in which farmers in the group are connected with external stakeholders such as experts, input firms and buyers, and so on. All transactions are performed through mKRISHI technology.



Phase I. Training, capacity building and data digitization



Phase II. End-to-end integration.

Figure 3. Phases of PRIDE Model Implementation

mKRISHI® Technology

The mKRISHI® – a patented Mobile Based Personalized Service Delivery Platform is the core technology used in the PRIDE model. This enables two-way data and information exchange between end-users such as farmers, field staff, and repositories of knowledge such as virtual knowledge banks, domain experts, input providers and procurement officers (PO). It requires professional and optimized management of resources, groupings of growers, provision of access to advisory or consultancy information, backward linkages (agricultural inputs and credit), forward (market) linkages, improving data visibility and enabling data analytics in a currently unorganized, unstructured sector. With this technology it is possible to effectively harness the power of farmer numbers under a common umbrella, coupled with a smooth flow of data and information to bring more structure into this sector (more on mKRISHI® is found in Appendix 2).

Implementation of Technology

The Farmers Producers Organization (FPO) and PRIDE envisaged using the viz. tomato for the purposes of the study since it is one of the major vegetables grown in the study region during the Kharif (monsoon) season (June–January).[1]

The Saptshrungi Farmer Producer Company Limited (SFPCL) associated with Agri Services Foundation (ASF) primarily works in Dindori tehsil of the Nashik district for the progress of farmers in the area. Prior to project implementation, farmers were facing problems related to crop productivity, availability of quality agricultural inputs and lack of access to better markets. To overcome these problems, SFPCL introduced the ICT-based PRIDE platform. Implementation started with Phase I involving awareness education regarding the technology and benefits, followed by digitization of data from the farmers and other stakeholders. Field staff associated with SFPCL trained the member farmers in technology and usage. Further field staff registered farmers on the mKRISHI system and digitized their unique profiles, including plot and crop specific details. In Phase 2, member farmers were provided personalized services at the field level and were connected with other stakeholders in the agricultural chain. These services are supplemented by occasional visits from field executives. Results of these actions are detailed in the following section.

Results and Implications

Prior to the technology implementation (years 2012–2013), farmers in the study region, were obtaining an average tomato yield of sixteen tons per acre. From 2013–2014 onwards, around 1140 farmers spread over thirty-seven villages have participated in the project. Initially they started with digitization of the farmer base, plots, crops followed by the creation of *crop protocols* (scientific crop management practices). These crop protocols are disseminated through *Interactive Voice Response* and supported by the *Agro-Advisory Module* which provides two-way communication between farmers and the agricultural experts. Farmers are further supported by occasional visits from field executives to farmers' fields. Timely *Alerts*

[1] Tomatoes were selected as the major crop in this study region. Projects in other regions focus on major commodities including cabbage, cotton, grapes, onions, pigeon pea, potatoes, soybeans, sugarcane. We found the challenges farmers face are similar in all geographies.

and Broadcasts on forecasted weather and associated risks are also dispersed to member farmers.

In 2014–2015, SFPCL had 2162 farmers encompassing 1661 acres. Growth in productivity of tomato crops from years, 2012–2013 to 2014–2015 is illustrated in Figure 5 and Table 7. It is observed that productivity of tomato crop increased from sixteen tons/acre in the year 2012–2013 to twenty-seven tons/acre in 2013–2014; and thirty-five tons/acre in 2014–2015. Similarly, the number of farmers participating in the project has increased from 586 in 2012–2013 to 1140 in 2013–2014; and 2162 in 2014–2015. Thus, due to the personalized crop protocol, agro-advisory and timely alerts, the average increase in productivity was found to be 64% in 2013–2014; and 112% in 2014–2015. It also contributed to around a 90% increase in farmer participation in the second year.

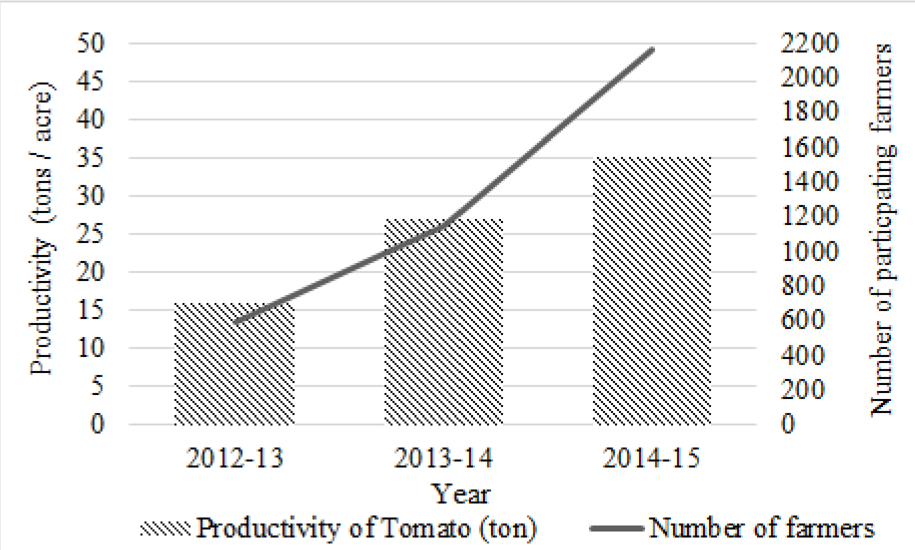


Figure 5. Growth in tomato productivity and participating farmers

Table 7. Growth in tomato area, productivity and number of participating farmers

Particulars\Year	2012–2013	2013–2014	2014–2015
Number of farmers	586	1140	2162
Area under Tomato (acre)	612	1010	1661
Productivity of Tomato (ton/acre)	16	27	35

The *Agricultural Input Management* module facilitates agricultural input requests from member farmers, aggregates the requests and communicates with the agricultural input providers. This results in larger group demand for agricultural inputs and associated optimizations on the costs incurred by farmers for inputs. It also helps input suppliers to plan their production and distribution activities accordingly. The member farmers are collectively ordering the agricultural inputs using this module and benefiting by purchase of quality inputs at reasonable prices at their door step. The outcome of this module combined with crop protocol, agro-advisory services, and weather alerts, resulted in optimizing the cost of production. Reduction in cost ~USD 227.38 per acre for tomato crop was observed in 2013–2014.

An essential implementation activity by SFPCL is the collective marketing of agricultural produce to agro processing companies and markets. This step is easy because all relevant data and information is stored within it. Farmers acquire the *market intelligence* from the system and accordingly plan when to harvest produce. Member farmers raise sell requests through the *Harvest Management Module* and communicate their plans for the harvest and sale of produce. Accordingly, SFPCL collects the harvested tomato from its member farmers, undertaking proper grading and sorting and finally marketing the aggregated produce to nearby processing industries and markets. Buyers registered on the system equally benefit as they receive requisite quantity and quality of produce directly from farmers. The SFPCL has established marketing agreements with two tomato processing companies in the area. Member farmers collectively market their produce at prevailing market rates, without any dependency on intermediaries. The system is also helpful for SFPCL in maintaining finance, audit and compliance at an organizational level, including bulk financial transactions.

Conclusion

Efficient data collection, management and its further use at the right time and with the right user base will help the agricultural community as a whole. The innovative PRIDE model has proven success in its pilot in Nashik district of Maharashtra with SFPCL. It helped farmers follow scientific crop management practices, receive correct and timely agricultural advice from experts, timely weather alerts to minimize associated risks, and collective demand for inputs and sale of the produce thereby increasing the returns from agriculture. The benefits of the model are not limited to the farmers only but to the other stakeholders involved in the agricultural chain which are agricultural input providers, agricultural processing companies, buyers, and so on. All this is possible because of mandatory data collection, effective organization, processing and efficient dissemination through the use of mKRISHI technology. Further it creates employment opportunities in rural areas (field staff, experts, and project managers).

This model is holistic, scalable and very promising for sustainable agriculture which leads to empowerment and growth of rural India. It caters mostly of the needs of various stakeholders involved in the agricultural namely farmers, service providers, researchers, extension workers, processers and market stakeholders.

References

Accenture Digital, 2015. Digital Agriculture: Improving Profitability. Website: https://www.accenture.com/_acnmedia/Accenture/Conversion-Assets/Dot Com/ Documents/Global/PDF/Digital_3/Accenture-Digital-Agriculture-Point-of-View.pdf. [accessed May, 2016].

Directorate of Economics and Statistics. 2013. *Pocket Book on Agricultural Statistics 2013*. Department of Agriculture, Cooperation and Farmers Welfare. Ministry of Agriculture and Farmers Welfare. Government of India. http://eands.dacnet.nic.in/ [accessed May, 2016].

DAC. 2014. Annual Report, 2013–2014. Department of Agriculture and Cooperation, Ministry of Agriculture, Government of India, Krishi Bhawan, New Delhi-110001. [accessed May, 2016].

e-agriculture. 2012. e-Sourcebook. Forum 3: ICT in Agriculture, ICT for Data Collection and Monitoring and Evaluation. June 2012. http://www.ictinagriculture.org/node/133 [accessed May, 2016].

Ghoge, Ketaki. 2013. Group farming may change the face of agriculture in state. *Hindustan Times*. Mumbai. 18-May. http://www.hindustantimes.com/mumbai/group-farming-may-change-the-face-of-agriculture-in-state/story-io5uqXyLgwaQhLHm7ttUGL. html. [accessed April 2015].

Government of Maharashtra. 2014. Gazetteers of Nashik District. https://cultural. maharashtra.gov.in/english/gazetteer/Nasik/index.htm. [accessed August 12, 2015].

Gupta, Seema and Parida, Ganesh. 2013. Group Farming: A community farming initiative. *LEISA India* 15.2: June. http://www.agriculturesnetwork.org/magazines/india/farmers-and-markets/group-farming. [accessed September 24, 2015].

ICT in Agriculture. 2012. Module 1: Introduction: ICT in Agricultural Development. http://www.ictinagriculture.org/sourcebook/module-1-introduction-ict-agricultural-development [accessed August, 2014].

IKSL. 2016. IFFCO Kisan Sanchar Limited. http://www.iksl.in/ [accessed March, 2016].

IITB. 2016. aAQUA: Almost All Questions Answered. http://aaqua.persistent.co.in/aaqua/forum/index. [accessed March, 2016].

IIIT. 2012. Overview and Advantages of eSagu Framework - An IT-based Personalized Agro-Advisory System, IIIT, Hyderabad and Media Lab Asia. http://agriculture.iiit.ac.in/ esagu2012/overview.php. [accessed March, 2015].

Kumar, P. S. K. 2011. Mapping and Preliminary Evaluation of ICT Applications Supporting Agricultural Development: An IFC sponsored Study in Uganda, India, and Indonesia. Presentation. ACDI/VOCA. http://acdivoca.org/sites/default/files/attach/legacy/site/Lookup/ACDIVOCA_MappingEvaluationofICTAppsforAg/$file/ACDIVOCA_MappingEvaluationofICTAppsforAg.pdf [accessed May, 2016].

LifeLines Agriculture. 2016. http://lifelines-india.net/agriculture. [accessed March, 2016].

MoA, GoI. 2015. *Year end Review for the Ministry of Agriculture for the Year 2014-15*. Press Information Bureau, Government of India, Ministry of Agriculture. 22-December, 2014. 15:28 IST. http://pib.nic.in/newsite/PrintRelease.aspx?relid=113870.

Pandit, Pagar Mansaram. 2012. Agricultural Development and Land use Pattern in Nashik District of Maharashtra, India. *Mediterranean Journal of Social Sciences* 3(16):151–161. December. doi:10.5901/mjss.2012.v3n16p151.

Planning Commission, GoI. 2015. Sector-wise contribution of GDP of India. Planning Commission, Government of India. 8-July. http://statisticstimes.com/economy/sectorwise-gdp-contribution-of-india.php. [accessed October, 2015].

RML. 2016. Reuters Market Light. http://rmlglobal.com/ [accessed on May, 2016].

Shodhganga, 2013. Chapter-II: Profile of Study Region. http://shodhganga.inflibnet.ac.in. [accessed April, 2015].

The Guardian. 2013. How ICT tools are improving efficiency of agricultural development. http://www.theguardian.com/global-development-professionals-network/2013/jan/24/data-collection-evaluation-technology-agriculture [accessed May, 2016].

The World Bank. 2011. Food Price Watch. Poverty Reduction and Equity Group, Poverty Reduction and Economic Management (PREM) Network. *The World Bank* February. http://siteresources.worldbank.org/INTPREMNET/Resources/Food_Price_Watch_Feb_2011_Final_Version.pdf. [accessed October, 2014].

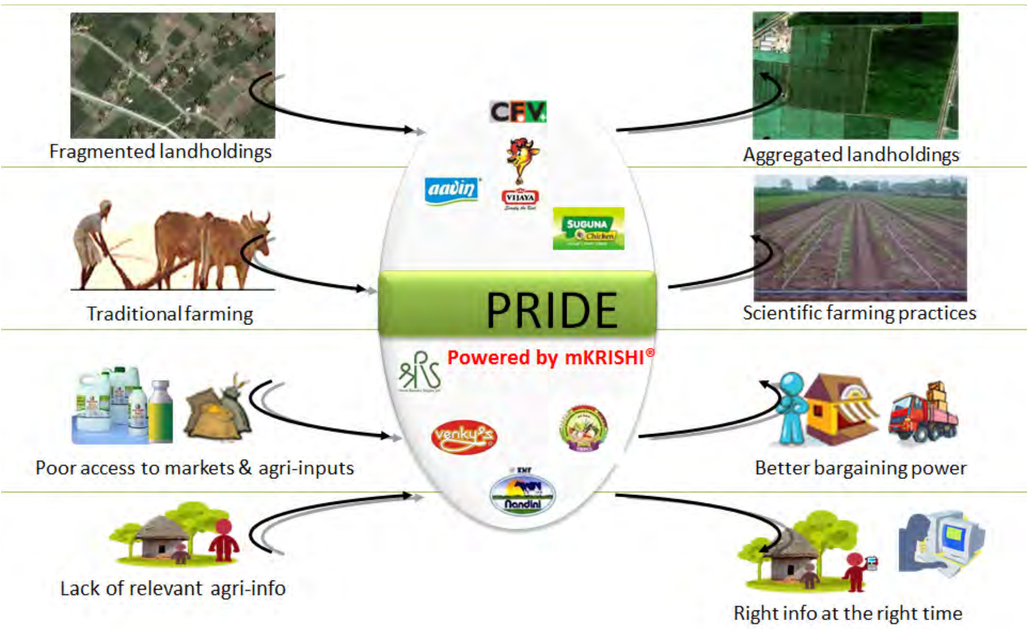
The World Bank. 2011. ICT in Agriculture e-Source Book, Connecting Smallholders to Knowledge, Networks, and Institutions, Report Number 64605. November. http://www.ictinagriculture.org/sites/ictinagriculture.org/files/final_book_ict_agriculture.pdf. [accessed May, 2016].

The World Bank. 2016. Overview. http://www.worldbank.org/en/topic/agriculture/overview. [accessed April, 2016].

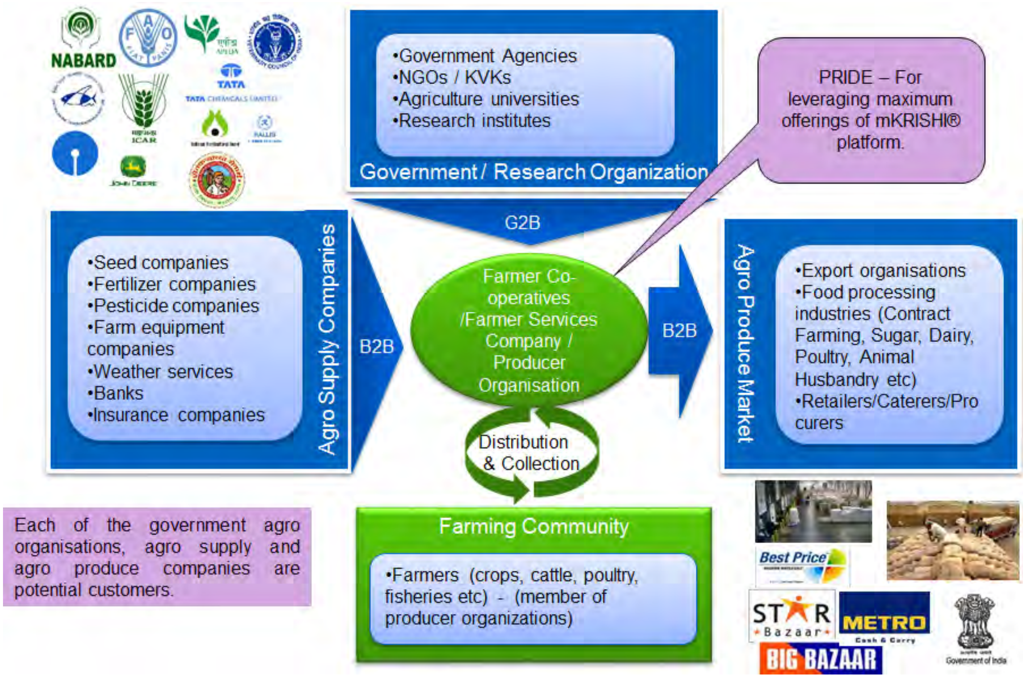
Vodafone Group. 2015. Connected Farming in India, How mobile can support farmers' livelihoods.https://www.vodafone.com/content/dam/sustainability/2015/pdf/connected -farmers.pdf. [accessed May, 2016].

Appendix 1

PRIDE™ is powered by Tata Consultancy Services' mKRISHI technology. The model is primarily based on the agricultural data and information collected from farmers' fields, agricultural universities and research organizations, weather data, nearby markets, inputs suppliers, etc. Data collected from farmers includes basic, financial, family and other socioeconomic details; plot details, property location, history; soil and water test results; crop details; etc. All collected data and information is processed and converted into meaningful agricultural advice, alerts, and broadcasts transmitted to the respective users.



(a) Bridging the Gap



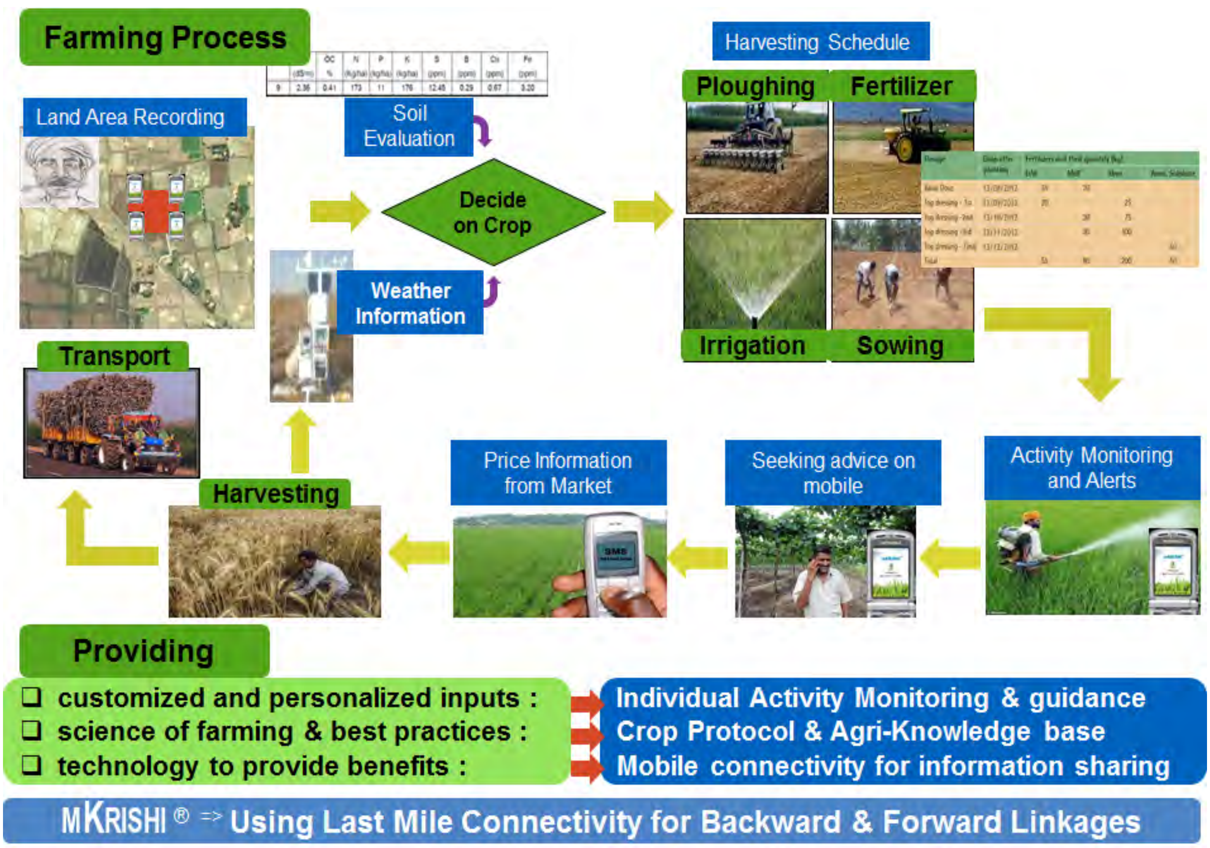
(b) Partners and collaborators

Appendix 2

Features of mKRISHI Technology

mKRISHI is a business solution combining the technologies of big data, Geographic Information Systems, analytics and mobile apps for enterprise management.

- Stakeholders registration and data management
- Agro-advisory
- Best practices
- Alert and broadcast services
- Weather forecast, reporting feature
- Agricultural input management
- Market intelligence
- Agricultural supply chain management services



International Food and Agribusiness Management Review
Special Issue - Volume 19 Issue A, 2016

A Blueprint for a Big Data Analytical Solution to Low Farmer Engagement with Financial Management

Cathal O'Donoghue[①] [a], Alistair McKinstry[b], Stuart Green[c], Reamonn Fealy[d], Kevin Heanue [e], Mary Ryan [f], Kevin Connolly [g], J.C. Desplat [h], Brendan Horan [i], and Paul Crosson[j]

[a] *Professor,* [c] *Research Officer,* [d] *Research Officer,* [e] *Research Officer ,*[f] *Specialist,* [g] *Specialist,* [i] *Research Officer* [j] *Research Officer, Teagasc, Oak Park, Carlow, Ireland*

[b] *Compputational Scientist,* [h] *Director, Irish Centre for High End Computing, National University of Ireland, Galway, Ireland*

Abstract

As the market environment for farming has become more complicated, the need for farmer engagement in financial management has increased. However, financial management decisions need to consider individual farm environmental conditions. This paper discusses the design of a new big-data based analytical solution for low farmer engagement in financial management—a Farm Financial Information System (FARMFIS). Using a pastoral based livestock system as the case study, the methodology required to develop this predictive Information System is described. Building upon real-time weather, satellite grass growth and soil information, a local setting and a bio-physical model of weather and market changes on farm level economic outcomes are utilized. The aim is to use the back-end framework described here to develop decision support tools for farmers to provide benchmark information in relation to the financial and technical attributes to a similar top, middle or bottom one-third performing farm. This information can help farmers engage more meaningfully in their own management decisions, technologies, and practices.

Keywords: big data, bio-economic modelling, decision support tools, financial management

[①]Corresponding author: Tel: + 353.87.1211489
Email: C. G. O'Donoghue: cathal.odonoghue@teagasc.ie

Introduction

The need for farmer engagement in financial management has increased as a result of greater complexity in the market environment for farming in terms of greater volatility, more complicated investment environments and viability challenges. At all levels of profitability improved financial management is required. However, farmers are more likely to take up agricultural technologies and practices than financial ones (Hennessy and Heanue 2012). Given this more complicated farm operating environment, there is a need for greater planning in order to provide greater resilience.

While much information is available to assist financial management and/or planning such as the eProfit Monitor (ePM) decision support tool (Morrow et al. 2004), take-up of such practices and technologies has been low. However, information provided by such decision support tools is essential for improved planning. Farmers using the ePM planning tool are ranked as top, middle and bottom performing farms on the basis of gross margin per hectare so that farms can benchmark their progress. This annual income measurement is strongly correlated with longer-term net profit (Teagasc 2015). While usage has increased substantially over time, only about 10,000 Irish farmers are using the eProfit Monitor, representing only a fraction of the population.

US research found that farmers “who conduct detailed financial analyses are substantially more profitable than the farmers who…did not make the calculations” (Gloy and LaDue 2003). Macken–Walsh et al. (2015) identify challenges concerning the current use of advisor managed interaction with existing decision support tools, where participation is often motivated solely by scheme incentivization, but without internalization of the information in their decision making. Further, Macken–Walsh et al. (2015) identifies that “potentially, the use of financial decision support tools may lead to ‘conscientization,' among farmers, where they come to realize the economic potential of their businesses and the potential of these tools. Dillon et al. (2008) report that 57% of Irish dairy farmers view financial management tools as time-consuming. Internationally, Gloy and LaDue (2003) found that although financial technologies were in use, they were often misunderstood and underutilized. Thus, it would appear that despite long-term benefits, farmers are reluctant to engage with financial and business planning because it is either too difficult to use or time-consuming to compile data, particularly relative to the financial return on investment on lower-income farms.

This lack of understanding and use of financial technologies is of concern within, for example, highly debt financed farm businesses where strict financial control and cash flow monitoring is essential. The net result is that farm-level financial management practices are not part of the routines of the farms’ operations, where routines are understood in the evolutionary economics sense to be ‘ways of doing and ways of determining what to do’ (Nelson and Winte 1982). Importantly, routines in a functional sense coordinate the other resources of the farm leading to their productive utilization (Dosi et al. 2000). Effectively, this means that financial tools are not part of many farmers’ management repertoires, although they need to be.

The Importance of the Environmental Context of Farms

Many livestock systems involve housing animals indoors for much of the year. For these systems, the environmental context of individual farms may not have a large impact on the economic success of the farm business. However, in pastoral (grass-based) livestock systems, environmental factors such as soil type, rainfall, soil temperature and soil moisture deficit, can have significant impacts on start and end dates of the grass-growing season, on the length of the grass-growing season, on grass yields and on soil trafficability. These impacts are compounded for farms that additionally grow their own supplementary feed requirements. All of these environmental factors vary across space, therefore, if financial management planning systems are to be meaningful, they must take this environmental and spatial variability into account.

Pastoral farming differs from most other businesses as it is context specific and spatial modelling requirements are different from other types of businesses. Thus, the modelling solution described in this paper is unique to the land-based farming context. The main technical challenges in predictive modelling of financial results are the spatial agronomic condition, the nature of farm system, including animal stocking rates and types, and the level of farm efficiency in terms of outputs and costs. Viewing spatial and government administrative data has provided solid grounding in the agronomic and system situation respectively. The remaining challenge is to model cost and production efficiency and farm subsidies, conditional upon the spatial and system situation.

A challenge then arises from the need to develop a Financial Information System (FIS) decision support tool that can give benchmark information in relation to financial and technical information that takes both the environmental context and the varying degrees of farmer engagement or skill into account. An additional challenge lies in delivering this information in a way that does not involve transaction costs that farmers perceive to be high, for example, completion of the ePM requires high-level data. Interestingly, in developing indicators of innovation on Irish tillage farms, Hennessy et al. (2013) report that while the adoption of innovative practices such as forward contracting and soil testing is highly correlated with economic performance, IT usage on farms is more widespread across farm economic performance. The motivation for farming is varied, from profitable commercial-minded farmers, to non-economically viable lifestyle farmers. Nevertheless, given the multi-faceted nature of farming and the increasingly complicated operating environment, it can be argued that it is necessary for farmers of all types to engage in more planning. While commercially focused farmers may already be more engaged in planning, using for example farm management accounts and existing decision support tools, less profitable farmers are less likely to engage. The greatest challenge, therefore, may be to provide information that is more accessible, allowing for differential farmer engagement (Oliver et al. 2012).

For lower income farmers to engage, the overhead of data collection and analysis needs to be lower than for existing decision support tools. Predictive approaches based on existing administrative and other real-time data sources can potentially allow for personalized information with lower overhead, which might enable greater usage and engagement. Of course the greater the reliance on predictive data than actual data, the lower the accuracy, but it is likely that some information, even if simulated, is better than no information.

In essence, there is a need to develop a predictive ePM that could provide simulated benchmark information for farmers in relation to the financial and technical attributes to a similar top, middle or bottom one-third performing farm. This modelling approach would counter the data-collection challenges faced by farmers in engaging with financial planning tools such as the ePM. The ability to additionally benchmark the environmental conditions of the farm would allow for a refining of the top, middle and bottom financial and technical benchmarks. This would help farmers better engage with the management decisions, technologies, and practices required for their specific spatial and environmental conditions.

Significant quantities of data are collected either for administrative purposes or utilizing remote sensing. Similarly, there are large complementary administrative spatial data assets available for use in this type of analysis. Much of the information is available to develop a predictive information system that can provide this benchmark system. However, the input of back-end statistical, spatial analysis, agricultural systems behavioral and ICT science is necessary to develop this capacity.

In order to allow farmers to engage more easily with the financial aspects of their business, it is necessary to understand the attributes of their enterprise at a local scale, with local specific agronomic drivers (such as soils, weather, altitude, etc.) and localized management decisions in relation to land base, system and stocking rate to:

- develop a bio-economic annual profit function based upon observed farm characteristics
- incorporate farm management decisions and resulting efficiency by understanding the technical and financial characteristics of top, middle and bottom farmers
- understand how to present complex financial and technical information to farmers.

This paper discusses the data elements and analytical components that form part of a blueprint design for a new Big Data based analytical solution, a Farm Financial Information System (FARMFIS), which facilitates easier engagement in financial management planning, taking individual farm locations and environmental contexts into account. We focus in particular on pastoral grass-based livestock farmers who face multiple complexities of managing the herd and a weather dependent grass crop as well as managing their interaction with the market in terms of inputs and outputs. Ireland’s mild maritime climate provides a competitive advantage in grass growth, making it the country most reliant on grass based livestock farming in the EU.

Figure 1 presents a flow diagram of the FARMFIS decision support tool. The first component is the bio-physical methodological framework for the farm financial information system. The most important time variant, agronomic driver of grass growth is weather. In order to understand the drivers of grass growth, the weather and soil parameters are extracted at grid points, equivalent to the remote sensing based grass growth measures. In order to understand the impact of differential agronomic conditions and grass growth across the country, it is necessary to link this data to farm data, management decisions and outcomes, which will then be linked to market prices to model the consequential market impact of the interaction between these bio-physical processes. Spatial microsimulation methods are utilized to create a base data set. The final stage of the system models the economic impact at farm level of biological systems on individual farms across the country at the spatial scale. To accomplish this, a bio-economic farm systems model is utilized. Our model builds on this approach by simulating economic outcomes related to animal demographics, feed supply, feed demand, imported feed, other costs and animal outputs on the spatially referenced farm and biophysical data to generate farm-level profits.

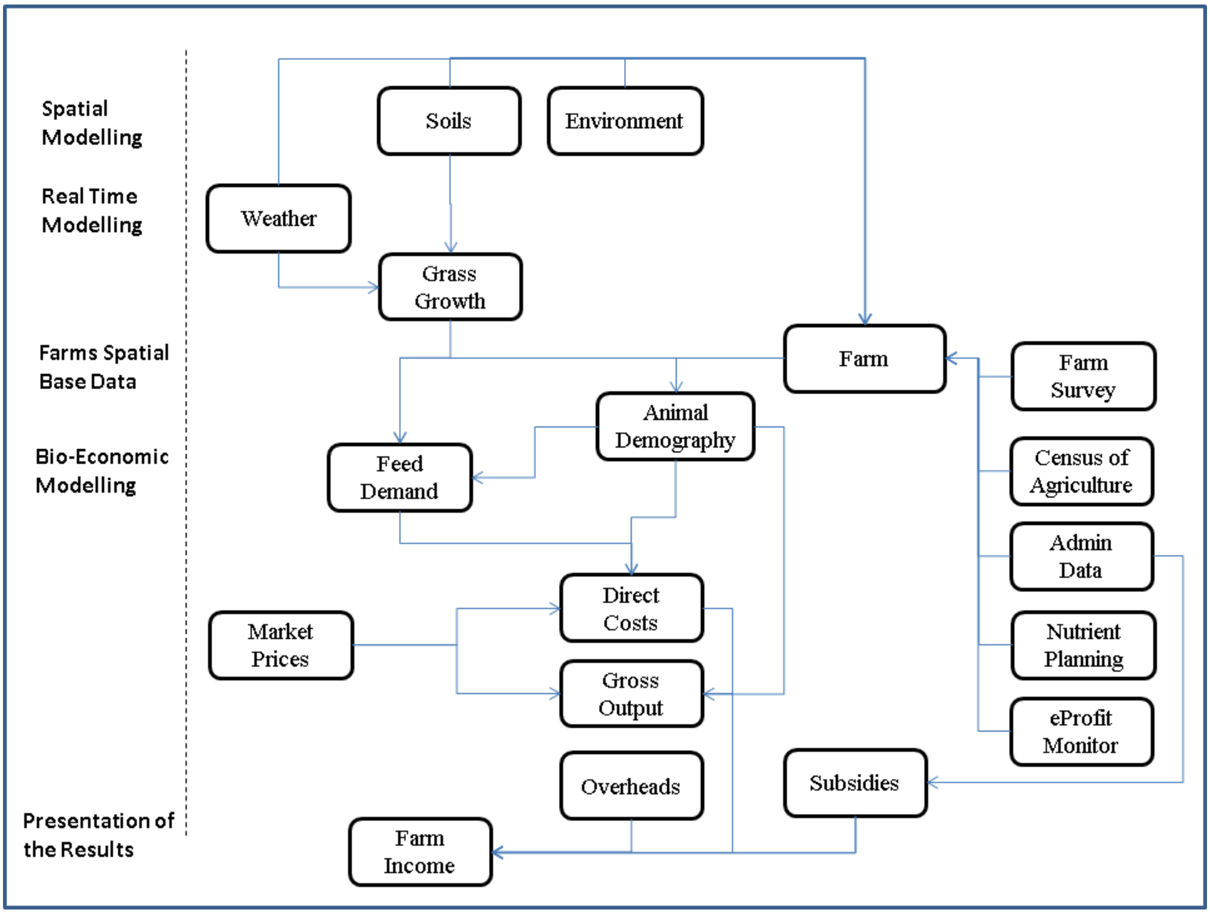


Figure 1. Structure of the FARMFIS Decision Support Tool

This paper describes a conceptual blue print for developing a decision support tool. After discussing the context in relation to the development of the model in section 2, various components of the blueprint are discussed. The bio-physical component is outlined in section 3, followed by the preparation of the base data and bio-economic system in section 4. The data requirements are charted in section 5, with a summary and next steps presented in section 6.

Extension Context

Cattle and Dairy Sector Context and Requirements of Financial Planning

Global demand for food is anticipated to increase 60% above current levels by 2020 (FAO 2015). At the same time, the increasingly international nature of food trade and associated trade policy disruptions have brought about unprecedented volatility in food prices which directly impact the financial performance of farm businesses (Shadbolt et al. 2013). Farming is widely acknowledged to be a financially risky occupation with an ever-changing landscape of possible price, yield and other outcomes that affect farm financial returns (Folke et al. 2002). Farm systems are complex and diverse, based on the resources which are unique to the farm, operating in volatile natural economic and policy circumstances. Such systems represent the collective response of farm businesses to remain viable and grow in the face of risks and uncertainties (Kaine and Tozer 2005). Due to an increasingly turbulent environment, recent studies suggest that financial evaluation of alternative farming systems must consider both the long term average profitability and the stability of farm income over time. The challenge for many farmers is to develop and implement farming systems with the preferred combination of activities and resources to mitigate these physical and financial risks and provide sustainable economic returns (Dillon et al. 2008).

For European Union (EU) dairy farmers, this challenge has been heightened in recent years due to a combination of reduced market supports and an associated increased exposure to more volatile global market prices coupled with reducing EU farm subsidies. As an export-oriented industry, the volatility of Irish dairy producer milk prices has increased four-fold during the last decade (Loughrey et al. 2015), and taken together with input price inflation, has resulted in increasingly volatile farm incomes. In an uncertain environment, improved farm financial management planning is a key attribute to helping farmers deal with future challenges and shocks (Mishra et al. 1999).

Beef production is the most widespread farm enterprise in Ireland accounting for almost 80% of the 139,000 farms in the national population and 34% of the gross output value from the agri-food sector. This output is largely generated from the suckler beef cow herd which comprises approximately half of the total number of breeding females, with the remainder originating from the dairy sector. Despite the significance of the beef sector, farm family incomes are low, with many farms operating at a loss when EU and national farm support payments are excluded. The Teagasc National Farm Survey (NFS) (Hanrahan et al. 2014), which is part of the Farm Accountancy Data Network in the EU, estimates that average farm income (including the EU direct payments and agri-environmental scheme subsidies) for suckler beef cow and beef finishing farms was €9,541 (US $12,526) and €15,667 (US $20,569) respectively, in 2013 (Hanrahan et al. 2014). The level of farm employment by the farmer and/or spouse on suckler beef and beef finishing farms is high at 56% and 47% respectively. Therefore, beef farms in Ireland are heavily reliant on EU payments, and alternative sources of income to support the farm family (Hanrahan et al. 2014).

Given the abundant availability of grazed grass as a low cost and high-quality ruminant animal feed, Irish suckler beef systems are predominantly pasture-based with the majority of cows calving in spring in order to match the onset of seasonal grass growth. The grass growing season ranges from approximately 250 days in the north-east to 330 days in the south-west with a yield difference of approximately eleven tons dry matter (t DM) per hectare, per year vs. fifteen t DM per year, respectively (Brereton 1995).

Existing Extension Support for Farm Financial Planning and Challenges

Individual farm financial appraisal and forward planning are built around initially conducting a benchmarking analysis of the farm performance at the whole farm as well as the enterprise level. The Teagasc eProfit Monitor (ePM) has built its reputation as the leading financial benchmarking analysis system available for extension services in Ireland. The ePM analysis is produced in the form of a standardized report of the farm financial output, expenses and profit for the most recently completed year of trading of the farm business. The ePM system utilizes available electronic sources for large amounts of the input data to increase the speed and accuracy of the data entry process.

Extension advisers guide farmers in the collation of the required input data but the main focus of the advisers is on ensuring the analysis is representative of actual farm financial performance and also in identifying the key efficiency decisions indicated by the analysis. As such, the analysis acts as a "health-check" to assess how the business is progressing. This helps to identify areas to concentrate on during subsequent planning and budgeting and to set the baseline for whole-farm forward financial planning.

The Teagasc Farm Business Monitoring system contains a planning / budgeting process that can be short term in the form of a one or two-year cash flow budget or alternatively a five-year financial projection can be completed to check on the long-term feasibility of planned change or investment projects. Key to the potential accuracy and credibility of this forward projection is to build on the actual farm performance ePM analysis. Robust projections for future farm output and input prices along with accurate modelling of changes in farm efficiency are also important, as the planned change is implemented and becomes imbedded in the normal running of the farm. This is particularly relevant in the case of planned investment involving scale increases or radical enterprise change. The development of a robust and validated model that can simulate possible stress-testing scenarios and greatly assist farmers and their extension advisers in assessing the risk elements associated with the proposed change.

Extension: Getting Farm Financially Fit

The development of the FARMFIS builds upon a multi-actor national extension program, "*Get Farm Financially Fit*[1]" which aims to improve financial management and business planning. A network of twenty-three extension, farming, financial, media and agribusiness organizations formed a network to have an impact in this area—recognizing the need for improved financial and business planning amongst farmers, while recognizing that (for a variety of reasons) the demand is not there. The network held a national campaign during 2015, which received significant media coverage, attracted over 1000 farmers to public meetings and has been followed by special supplements and a series of fortnightly *Get Farm Financially Fit* articles in the specialist farming media. The concept of the Financial Information System (FIS) is seen by the network as a vehicle that can assist in the campaign for improved financial and business planning.

To understand how the extension processes and activities outlined above will need to be channeled in the context of FARMFIS, it is useful to be guided by Feldman and Pentland (2003) and Pentland and Feldman (2005), understanding of routines which isolate their ostensive, performative and artefact aspects. Specifically, by identifying the performative aspects of farm financial management routines (the practices/tools that farmers actually use and how they make decisions about farm finances) and their attitudes towards such practices and decisions, extension advisers will be able to clarify how a new artefact (inputting data into, and the use of, FARMFIS) can be integrated with existing farm-level routines and knowledge, and ultimately change those routines.

Macken-Walsh et al. (2015) outline what we know from the literature and from advisory expertise in Teagasc, which extension methods and approaches have been successful in relation to understanding and influencing the performative aspects of routines around ePM. These need to be applied to the FARMFIS case and include: actively generating farmers' perceptions that FARMFIS is useful and relevant; facilitating farmers' understanding of how FARMFIS works; accommodating different levels of farmer competence; involving the spouse and other family members to increase the impact of FARMFIS on farm-level

[1]Get Farm Financially Fit: http://www.teagasc.ie/rural-economy/farm-financial-fitness/.

decision-making: building esteem and pride around the use of FARMFIS and awareness of farmers' financial sensitivities.

It is hoped that the inclusion of spatial and environmental data in the FARMFIS will improve farmer engagement in financial planning as it will allow extension initiatives such as the *Get Farm Financially Fit programme*, to present financial benchmarking information which takes specific environmental challenges into account, making it easier for farmers operating in challenging environmental conditions to have realistic and achievable benchmarks.

Bio-Physical Methodological Framework for the Farm Financial Information System

The main technical challenge in predictive modelling of financial results is the spatial agronomic condition, the nature of farm system including animal stocking rates and types, and the level of farm efficiency in terms of outputs and costs. We know from spatial and government administrative data the agronomic and system situation respectively. The remaining challenge is to model cost and production efficiency and farm subsidies, conditional upon the spatial and system situation. High quality, nationally representative survey data has enabled our teams to bridge this technical challenge.

Measurement of Grass Growth Using Remote Sensing

As grass (either as pasture or winter feed) is the main feedstuff of the temperate Atlantic dairy producing nations in Europe, it is important to understand grass production variability both spatially and temporally. From an international perspective, a recent report (CSIRO Australia 2014) from Group on Earth Observation (GEO), a global body of which Ireland and EU are members) states that: "*Currently there is no comprehensive global effort for monitoring the status and productivity of pastures and rangelands.*"

Globally, estimates of tillage crop yields from satellites are common, but grass yields are less common although they are being addressed in Australia and New Zealand. An important difference between grassland yields and crop yields is that final yield information is much less important in grass than estimates of current yield. Current systems only estimate grass conditions at relatively coarse spatial dimensions (sub-regional levels) in open rangeland systems, but there has been a very recent increase in research in yield monitoring in closed paddock scale operational systems (Dusseux et al. 2015; Stafford et al. 2013). In Ireland, accurate data on actual grass yields are limited to a few sites around the country and are published as growth rates as opposed to quantities of biomass. This is partially resolved by the new PastureBase Ireland grass growth recording system (Griffith et al. 2014), which is a spatially enabled database of 300+ farmers recording weekly growth rate measurements on their farms.

Satellite systems capturing daily images of the country allow us to expand from 300 farms to propose a seamless national, per-hectare coverage of weekly growth levels in Irish pastures. The system is potentially deployable in all grass-based dairy producing regions in temperate northern Europe.

This system builds on recent work in the Spatial Analysis Group of Teagasc in both grassland monitoring through machine learning (Barrett et al. 2014) and site-specific modelling relating grass biomass to satellite data using time series sensor based data NDVI (Normalized

Difference Vegetation Index) from the Moderate Resolution Imaging Spectroradiometer (MODIS), flown on two NASA spacecrafts (Ali et al. 2014). The main goal of this system is to use Pasture Base ground-truthing and the new Sentinel 2 satellite data from ESA in conjunction with Landsat 8 data from NASA[2], within a machine learning environment, to give weekly estimates of current biomass and total annual grass production at local farm and national levels. These satellite sources along with the active radar satellite Sentinel 1 will also be used to characterize management (grazing, silage, and hay harvesting) at a parcel scale.

Downscaling of Meteorological Data

The most important time-varying agronomic driver of grass growth is weather. The Irish Meteorological service (Met Éireann) collects daily data at climatological, synoptic and rainfall stations. Operationally this data is then assimilated into the Harmonie Numerical Weather Prediction (NWP) model (Seity et al. 2011; Van der Plas et al. 2012) developed by HIRLAM[3] consortium .

Harmonie is run at 2.5 km resolution four times daily, assimilating ground and remote observations. In addition, Met Éireann is currently completing a re-analysis of Irish weather data for the period from 1980–2014. Re-analyses are model "forecasts" that assimilate all the available observations over the period, thus producing a consistent gridded historical weather dataset. In order to understand the drivers of grass growth, the weather and soil parameters from the NWP model will be compared to the remote sensing based grass growth measures.

Understanding the Agronomic Drivers of Grass Growth

The current *Pasturebase Ireland* analysis (Griffith et al. 2014), based on farmer collected grass management data in Ireland, suggests that pasture performance (growth rate, accumulation, growing season, etc.) variation is much wider over smaller scales than existing models suggest. Current agronomic studies include fixed effects of environment implicitly rather than explicitly, and analyze climate rather than weather. This allows for the coarse capture of general agronomic performance as a function of location but does not allow for a detailed understanding of the interaction of management, environment and weather in farm performance. By explicitly building a spatial and temporal model of environmental and weather impacts will enable better understanding of issues such as risk (e.g. how exposed are farm systems to bad weather) and agronomic potential.

The next stage in the analysis is to understand the spatial drivers of grass growth. The dependent variable is based on the grass growth data measured using remote sensing. Based on earlier work from a single site (Hurtado-Uria et al. 2013) this study analyzes grass growth across on a varying spatial and temporal continuum. Inputs to the model include:

- Spatially varying soils data from the new Irish Soil Information System (Creamer et al. 2014).
- Spatially varying slope and altitude data.

[2] National Aeronautics and Space Administration
[3] The international research programme HIRLAM (High Resolution Limited Area Model) is a research cooperation of European meteorological institutes. The aim of the HIRLAM programme is to develop and maintain a numerical short-range weather forecasting system for operational use by the participating meteorological institutes.

- Time and spatially varying weather data.
- Spatially varying farm management data (livestock density) taken from the Census of Agriculture and potentially extended to time and spatially varying livestock density data taken from Administrative Data.
- Farm systems data from spatially enabled NFS.
- The model will combine dependent and explanatory variables using a spatial panel data statistical model to understand the relationship. Panel data analysis, with time-series satellite data treated as cross-sectional data at the pixel level, is the analytical bridge between remote sensing and agronomy approaches. Spatial (Baylis et al. 2011) and Geographical Weighted Panel data (Cai et al. 2012) approaches will also be of particular importance. Methods will be expanded from the area based, neighborhood (spatial lag) analysis approaches encountered in the literature to work with the available point and continuous surface data. The use of spatial panel data is a new and still developing analytical area and its combination with remote sensing data provides a novel approach which will make a significant contribution to the existing literature.
- One of the challenges of using some of these data sources is that the variables are not necessarily collected at the same spatial scale. For example, both the Soil Information System and the meteorological data are modelled, but spatially mapped surfaces are at different resolutions. This may have implications for determining the spatial resolutions that are most appropriate for the analysis. The explanatory power of the model will be tested at different spatial resolutions and develop estimates of the confidence intervals for this geo-statistical model at these spatial resolutions.
- Through the application of the Spatial Panel Data Analysis and Geographically Weighted Panel Regression methods, our ultimate aim is to develop a spatial agri-econometric model linking biomass accumulation to biomass utilization and to test scale dependency in the model. The model framework will develop.
- Develop a spatial model which includes environmental and management factors that explain temporal and seasonal variation in pasture growth performance in Ireland.
- a high-resolution map of pasture performance zonation in Ireland – identifying areas of potential under-performance as a result of prevalent environmental conditions.

Incorporating Real-Time Data Changes to Model Grass Growth: Data Assimilation

The availability of real-time data for meteorological components and for grass growth allows for real-time validation and improvement of the models using the data assimilation techniques used in meteorological forecasting. We propose to improve the real-time predictive capacity of farm specific annual grass and net grass supply models using data assimilation in the land and grass models.

Harmonie contains a land surface model, the SURFEX model (Le Moigne et al. 2009). This provides temperature, evapotranspiration and surface roughness parameters to the atmosphere component. Hence, this model contains a live model of the soil temperature and moisture characteristics of Ireland; to date, this information has not been evaluated. It is planned to progressively implement an ensemble-based data assimilation into a coupled land model, adding in-situ and remote observations of soil and vegetation to the model(s) and generating an ensemble of weather outcomes, which can be used to investigate variability in the grass growth model at the farm level.

An ensemble of initial states for Harmonie based on soil moisture, temperature, and agronomic characteristics will be tested to determine the effect on variability, and compared to observations. During the ensemble modes, the model runs with slightly different initial conditions are used to capture the variability and predictive skill of the forecast (Iversen et al. 2011).

The SURFEX surface model uses the ISBA soil model to provide soil temperature and moisture. This describes soil in terms of simple sand and clay fractions and soil depth. In addition, tree height and Leaf Area Index (LAI) are provided.

Currently, land observations are not assimilated by the Irish meteorological service; the soil state is driven by meteorological fluxes at the surface only. Assimilation will be enabled and tested based on observations from SMOS (Mahouf and Balsalmo 2015), ASCAT (Barbu, 2014); and Sentinel-3 (Lewis et al. 2012). Soil temperature and moisture at selected stations will be used to validate the soil state.

The Soil map of Ireland (Simó et al. 2014) provides detailed information on the soil composition. This will be used in the initial ensemble vectors. Similarly, the vegetation state in SURFEX is currently based on historical averages from ECOCLIMAPII (Seity 2011). This can instead be updated directly from the grass model and remote observations.

SURFEX currently has two assimilation methods, Optimal Interpolation (OI) and Extended Kalman Filter (EKF) (Duerinckx et al. 2012). Modern developments are typically based on Ensemble methods such as Ensemble Kalman Filtering (EnKF) (Evensen 2003). Such methods are preferable as coupled models become more complex (such as where grass models are driven by weather models). Here, EnKF or Bayesian Model Averaging may be applied. BMA has been used with Harmonie within the GlamEPS project (Iversen et al. 2011) which included Met Éireann and ICHEC and for high-resolution wind forecasting (Peters et al. 2013). EnKF is also suitable for land data assimilation (Zhou et al. 2006).

Linking Bio-physical Processes and Farm Data using Microsimulation and Farm System Bio-Economic Modelling

Microsimulation Modelling of Base Dataset

In order to understand the impact of differential agronomic conditions and grass growth across the country, it is necessary to link this data to farm systems, farm size, and animal demographics. In a subsequent step, this information will be linked to farm management decisions and outcomes, which will then be linked to market prices to model the consequential market impact of the interaction between these bio-physical processes.

In a fully operational decision support tool, we would utilize actual farm data (taken from administrative registers) along with spatial data and remote sensing, to provide simulated benchmark data for specific farms. However, there remain a number of challenges to achieving this in relation to accessibility and data cleaning. Therefore, we use synthetically generated representative data using data enhancement methods to create a synthetic spatial farm dataset (see O'Donoghue et al. 2013).

Because we require individual financial data, we cannot use small area analysis for this purpose (Ghosh and Rao 1994). Therefore, we require a method that maintains both spatial variability and micro-level variability such as spatial microsimulation (Clarke 1996). There is extensive literature described in O'Donoghue et al. (2014) covering many different policy areas, utilizing methodologies described in Hermes and Poulsen (2012).

In determining the methodology to use for the creation of a farm level spatial microsimulation model, we face a number of issues. While Iterative Proportional Fitting (Deming and Stephen 1940) could potentially be used to produce small area weights, it struggles to deal with the issue of heterogeneous stocking rates. Similarly, given how many districts have small numbers of farms in Ireland, the Deterministic Reweighting method (see Tanton et al. 2011) is potentially challenging. Simulated Annealing (Williamson et al. 1998; Ballas and Clarke 2000) was used to generate an earlier version of the model (Hynes et al. 2009) but has significant computational costs and also struggles with the spatially heterogeneous stocking rates.

Thus, we will use a methodology that is sample-based in order to (a) avoid the income smoothing concern of the weighting methodology; (b) be computationally efficient, and; (c) adjusted to improve the spatial heterogeneity of stocking rates. We utilize a method developed by Farrell et al. (2013) known as Quota Sampling (QS) which is a probabilistic reweighting methodology, whereby survey data are reweighted according to key constraining totals for each small area.

In this analysis, the farm-level survey data (NFS) is statistically combined with spatial Census of Agriculture data. The most recent Census of Agriculture was collected in 2010 and combined this with the Teagasc National Farm Survey (Hanrahan et al. 2014).

Bio-Economic Systems Modelling

Once estimates of bio-physical drivers of income are modelled at point scale, we will then need to understand how this affects the on-farm profitability at those points. For this, a bio-economic modelling system is required that links these characteristics to financial outcomes across a range of farms within their spatial agronomic context. Bio-economic systems models facilitate the integration and synthesis of knowledge from many areas of research including animal growth, grass growth, feed utilisation and farm management. In the context of the present study, the bio-economic systems model combines fluctuations in grass availability, farm-level characteristics such as animal demographics and ensuing stocking rates, which are the principal drivers (together with input and output price volatility estimates) with which to generate farm profit fluctuations.

Thus, at the core of the modelling system will be a bio-economic farm systems model that models the biological processes on farms of a particular type, with agronomic and grass growth conditions taken from the spatial weather and grass growth models and relates financial outcomes to biological processes across a range of heterogeneous farms.

At present, the models of this type used by the authors are single farm models and are based either on experimental farm data or utilise the characteristics of "average" survey farms (Crosson et al. 2006). Most farm systems models utilize typical farm data based upon experimental conditions (Doole and Romera 2013; Doole et al. 2013; Chardon et al. 2012). These models have been used to simulate the impact of changes in farm practices and technological adoption.

A significant agricultural systems research literature exists which analyses the components and relationships of the whole farm system to elucidate performance outcomes associated with both “endogenous” and “exogenous” activities to the system (Gordon 1969). A range of descriptions and applications of systems models have been published, many of which can be classified as mathematical programming models of production (Janssen and van Ittersum 2007).

Most models are based on hypothetical or representative farm types (Crosson et al. 2006; Wallace and Moss 2002) and have been typically developed for specific applications or locations. Whilst there are notable examples (such as Rotz et al. (2005) whose model includes weather and soil effect), few models have been developed to address multiple assessment areas or geographic locations, i.e. employ a generic framework or are designed to enable upscale of results to higher systems level such as national scale. Examples of models which have employed such generic frameworks to model farming systems for a variety of research questions include the German MODAM model (Kächele and Dabbert 2002; Zander 2001), the Australian MIDAS model (Pannell 1999), the European FSSIM model (Louhichi et al. 2010a, 2010b), and the Scottish ScotFarm Model (Shrestha et al. 2014). However these models are typically designed to model representative farm types based on a specific typology defined by some combination set of farm size, production intensity, production system (dairy, sheep, beef, etc.), biophysical descriptors, etc., in order to analyse grouped farms with similar characteristics in specific regional or agronomic zones.

There is thus a scientific gap in being able to model the impact of management and technological characteristics across a range of actual farms. In Ireland, Shrestha et al. (2014) developed a relatively simple bio-economic systems model utilizing typical farms on a regional basis with a simpler production system than the Moorepark Dairy Systems Model. A similar methodology has been employed in Scotland by SRUC (Barnes et al. 2014). At a European scale the European-wide equivalent dataset to the Teagasc National Farm Survey, the Farm Accountancy Data Network has been employed to develop systems models at a disaggregated scale (Janssen et al. 2010; Louhichi et al. 2010a, 2010b; Van Ittersum et al. 2009) using geo-referenced data (Green and O’Donoghue 2013).

However, both types of models could be criticised for having less realistic bio-economic systems than the single farm systems model. In particular, they lack the capacity to relate farm level outcomes to localised environmental and weather conditions and do not incorporate grass supply, which is one of the primary determinants of purchased feed for animal-based systems.

Simulation

To develop the base dataset of farm characteristics on which simulation will be based, we will utilize spatial microsimulation techniques as follows:

- The quota sampling generates the spatial distribution of farm size, farm system, and soil type, but does not incorporate localised agronomic characteristics such as weather, altitude.
- In order to make these data consistent with agronomic and grass growth data, we will estimate statistical models of the animal demographic, output and cost dependent variables in the Teagasc National Farm Survey as a function of farm and spatial characteristics (geo-referenced cost and production functions). This utilizes

geo-referencing of NFS linked to agronomic and environmental characteristics at the location of the farms.

- Utilizing the estimated statistical models, we can adjust the dependent variables using microsimulation to account for the localised agronomic characteristics. While localised ex-post calibration has been undertaken in the literature using alignment or calibration (Li and O’Donoghue 2014), agronomic based ex-post adjustment has not yet been used in the literature due to the unavailability of suitably geo-referenced micro data. The methodology developed here will extend the literature enabling these models to be used for more spatially disaggregated analyses such as the interaction between farming and water quality.
- In order to be able to create a localized farm financial information system, we will eventually develop a heterogeneous farm system model for dairy, cattle and sheep. However, the framework will initially be piloted for simpler sheep systems. This model will take as input the agronomic, grass growth, system, and animal demographic characteristics of the farm. Specific model components will be generated including the following modules: Animal type specific nutrition requirements; Feed Demand; Other Inputs; Farm Output; Market price and profit module linking volume inputs and outputs to prices utilizing methodology used in Shalloo et al. (2004) and Crosson et al. (2006), however applied to heterogeneous data. For annual income profit analyses, price projections from Teagasc Agricultural Outlook modelling will be used. Subsidies will be treated exogenously, given decoupling of CAP payments.

We allow for differential farmer engagement so farmers could access (top, middle and bottom) benchmark information for a farm with their agronomic characteristics, size, stocking rate, and system). Other farmers who interact with systems such as the eProfit Monitor could avail of greater detail as to their relative efficiency. To do this, we will simulate various versions of the model with different levels of actual and simulated data and compare it against raw data. As the simulation process and system are stochastic, we will use Monte Carlo simulations with different random numbers to develop confidence intervals for different farms.

Data

In a fully operational decision support tool, we would utilize actual farm data taken from administrative registers, spatial data, and remote sensing to provide simulated benchmark data for specific farms. However due to accessibility and data cleaning issues, we need to utilize an alternative data source in order to develop the FARMFIS model.

The CSO Census of Agriculture contains the spatial distribution of farms by the system, farm size, animal numbers, etc. In many ways, it contains similar data to that available on administrative registers. Similarly, detailed farm level data is available through the Teagasc National Farm Survey (NFS). Although these data have recently been geo-referenced and can be linked to spatial agronomic conditions, with a sample of about 1000 farms, the sample size is not sufficiently large to be able to undertake spatially representative analyses.
Therefore, as in the case of other spatially specific analyses, we will use synthetically generated representative data using data enhancement methods to create a synthetic spatial farm dataset, combining the best of both farm-level survey data and spatially disaggregated Census of Agriculture data (See O’Donoghue et al. 2013).

Scoping the Use of Administrative Data

In order to operationalize the Farm Financial Information System, we utilise existing administrative data sources and large complementary remote sensing spatial data assets. Examples of existing data include:

- Animal movement data are recorded on the Department of Agriculture Food and the Marine (DAFM) Animal Movement and Identification System (AIMS) system;
- Land use and land area on the Land Parcel Information (LPIS) system;
- Farm characteristics in the CSO Census of Agriculture;
- Farm subsidies on various DAFM administrative registers;
- Soils data in the Teagasc Soil Information System and the Soil Sample database;
- Agronomic and environmental data on Teagasc and EPA Spatial databases;
- Meteorological data from the Irish Meteorological Service ground stations; grass growth through satellite-based remote sensing;
- Fertiliser use in the Teagasc Nutrient Management Planning Software;
- Detailed farm activity data in the Teagasc National Farm Survey Database; and
- Farm financial data on the Teagasc eProfit Monitor system.

It is evident that much of the data needed to develop a predictive information system that can provide this benchmark system already exists. However, the back-end statistical, spatial analysis, agricultural systems, behavioral and ICT science needs further work to develop the capacity to process this data. Importantly, an appreciation of the potential use of integrating big data sources does not yet exist.

Summary and Next Steps

In this paper the development of a blueprint is described for a modelling framework to develop a Farm Financial Information System (FARMFIS) to assist farmer financial decision making, which will builds upon existing big data resulting from administrative, remote sensing, meteorological and survey data and a variety of different model methodologies to produce localized farm information.

Farmers who utilize FARMFIS could improve their financial and environmental performance by:

- improving their cost management through benchmarking against technically more efficient peers;
- making decisions about appropriate animal stocking rates that can improve both financial and environmental performance;
- adopting appropriate farm systems (dairy, cattle, sheep, tillage) appropriate to their capacity, land and financial needs;
- improving nutrient management; and
- making better investment decisions.

Real-time and predictive information about feed requirements and availability over the course of the year at a localized level would assist extension advisors to provide targeted local advice to provide early warning systems during difficult times. For example, national media coverage of the Fodder Crisis in 2013 gave the impression that the impact was nationwide, whereas in reality according to Teagasc remote sense information, the impact was more localized. Additionally, this information would allow for a prioritization of resources on a spatial basis. It would also improve the capacity of advisors to provide localized agronomic and planning advice to farmers.

Improved localized agronomic financial information can allow for:

- improved estimates of Soil Moisture by the Irish meteorological service;
- a better understanding of the impact of improved financial management by policy makers within the Department of Agriculture Food and the Marine; and
- assisting in the dissemination of key financial planning messages across a range of financial, agri-business and training stakeholder partners in the Getting Farm Financially Fit agri-sector network.

This paper provides a framework to make it easier to provide predictive financial information, drawing on a wide variety of big data sources and current financial and economic modelling techniques in the agricultural setting. However, modelling capability and data are not sufficient for the system to have an impact on facilitating decision making by farmers. The process by which farmers engage with financial data and make financial-related decisions is highly complex and crucially involves mediating farmer behavior (Macken-Walsh et al. 2015).

The experience of the agricultural extension experts in the project team is critical to maximizing impact as they provide an understanding of how farmers engage with this information and how they make consequential decisions. The prototype described in this paper forms the back-end analytical solution. In order to fully engage farmers, it requires co-designing with farmers on the front-end and interpreting the predictive financial data in a way that is meaningful to farmers.

In order to maximize the effectiveness of the approach, it will be necessary to provide outputs from the decision support tool in ways that are accessible to farmers with different technical skills. These will vary from online interactive tools for farmers with the need to access more detailed information, to simpler dissemination tools utilizing smart phones and support materials, as well as more general dissemination through farm media.

The contribution of this paper to the literature and novelty of the approach lies in the fact that while big data has been used extensively in precision agriculture in terms of agronomic decisions, real-time decision tools that focus on predictive full-farm financial benchmarking, utilizing real-time administrative and satellite data are new. The paper describes the conceptual blueprint that is being utilized by a team of Ph.D. students, agricultural extension specialists, agricultural economists and spatial analysts in developing a functional back-end system. Given the complexity of the modelling framework used as part of the decision support tool, the purpose of this paper is to outline the methodology in advance of the operational implementation of the framework.

References

Ali, I., F. Cawkwell, S. Green and N. Dwyer. 2014. Application of statistical and machine learning models for grassland yield estimation based on a hypertemporal satellite remote sensing time series. Paper presented at the annual meeting of Geoscience and Remote Sensing Symposium (IGARSS) July 13–18. http://iffimaan.com/projects/work/igarss2014.pdf. doi: 10.1109/IGARSS.2014. 6947634.

Ballas, D. and G.P. Clarke. 2000. GIS and microsimulation for local labour market policy analysis. *Computers, Environment and Urban Systems* 24:305–30.

Barbu, A. L., J-C. Calvet, J-F. Mahfouf, and S. Lafont. 2014. Integrating ASCAT surface soil moisture and GEOV1 leaf area index into the SURFEX modelling platform: a land data assimilation application over France. *Hydrology and Earth System Sciences* 18(1):173–192.

Barnes, A., S. Shrestha, S. Thomson, L. Toma, K. Mathews, and L.A. Sutherland. 2014. Comparing visions for CAP reforms post 2015: Farmer intentions and farm bio-economic modelling. *FACCE MACSUR Reports* 3(1): 3–2.

Barrett, B., I. Nitze, S. Green and F. Cawkwell. 2014. Assessment of multi-temporal, multi-sensor radar and ancillary spatial data for grasslands monitoring in Ireland using machine learning approaches. *Remote Sensing of Environment* 152(September): 109–124.

Baylis, K., N.D. Paulson, and G. Piras. 2011. Spatial Approaches to Panel Data in Agricultural Economics: A Climate Change Application. *Journal of Agricultural and Applied Economics* 43(03): 325–338.

Brereton, A.J. 1995. Regional and year to year variation in production. In: *Irish grasslands – their biology and management.* 12–22. Edited by D.W. Jeffrey, M.B. Jones, and J.H. McAdam. Dublin Royal Irish Academy.

Cai, R., D. Yu, and M. Oppenheimer. 2012. Estimating the effects of weather variations on corn yields using geographically weighted panel regression. Unpublished, Manuscript.

Chardon, X., C. Rigolot, C. Baratte, S. Espagnol, C. Raison, R. Martin-Clouaire and P. Faverdin. 2012. MELODIE: a whole-farm model to study the dynamics of nutrients in dairy and pig farms with crops. *Animal* 6(10):1711–1721.

Clarke G.P., 1996. Microsimulation for Urban and Regional Policy Analysis, London: Pion.

Creamer, R., R. Fealy, S. Hallett, J. Hannam, N. Holden, B. Jones, T. Mayr, I. Simo, and R. Schulte. 2014. Irish Soil Information System. Summary of Findings. EPA Research Report 130. EPA, Wexford.

Crosson, P., P. O'Kiely, F.P. O'Mara and M. Wallace. 2006. The development of a mathematical model to investigate Irish beef production systems. *Agricultural Systems* 89(2):349–370.

CSIRO Australia, 2014. Summary Brief: Establishment of GEO Global Rangelands and Pasture Productivity (RAPP), Group on Earth Observations.

Deming, W.E. and F. F. Stephan. 1940. On a least square adjustment of a sampled frequency table when the expected marginal totals are known. *Annals of Mathematical Statistics* 11(4):427–444. December.

Dillon, P., T. Hennessy, L. Shalloo, F. Thorne, and B. Horan. 2008. Future outlook for the Irish dairy industry: as study of international competitiveness, influence of international trade reform and requirement for change. *International Journal of Dairy Technology* 61(1): 6–29. doi: 10.1111/j.1471-0307.2008.00374.x.

Doole, G. J., and A. J. Romera. 2013. Detailed description of grazing systems using nonlinear optimisation methods: a model of a pasture-based New Zealand dairy farm. *Agricultural Systems* 122: 33–41. doi:10.1016/j.agsy.2013.08.001.

Doole, G. J., Romera, A. J., & Adler, A. A. 2013. An optimization model of a New Zealand dairy farm. *Journal of Dairy Science* 96(4): 2147–2160. doi:10.3168/jds.2012.5488.

Dosi, G., R.R. Nelson and S.G.Winter. 2000. Introduction, in *The Nature and Dynamics of Organizational Capabilities.* New York: Oxford University Press.

Duerinckx, A., R. Hamdi, J.-F. Mahfouf, and P. Termonia. 2015. Study of the Jacobian of an Extended Kalman Filter for Soil Analysis in SURFEXv5. *Geoscientific Model Development* 8:845–863 doi:10.5194/gmd-8-845-2015.

Dusseux, P., L. Hubert-Moy, T. Corpetti and F. Vertès. 2015. Evaluation of SPOT imagery for the estimation of grassland biomass. *International Journal of Applied Earth Observations and Geoinformation* 38: 72–77.doi: 10.1016/j.jag.2014.12.003.

Evensen, G. 2003. The Ensemble Kalman Filter: Theoretical formulation and practical implementation. *Ocean Dynamics* 53(4):343–367.

Farrell N., C. O'Donoghue, K. Morrissey, J. Lennon, D. Ballas, G. Clarke and, S. Hynes. 2013. The SMILE model: construction and calibration, in *Spatial Microsimulation for Rural Policy Analysis*. Springer-Verlag - Advances in Spatial Science.

FAO 2015. OECD-FAO Agricultural Outlook 2015–2024. http://www.fao.org/3/a-i4738e.pdf.

Feldman, M.S. and Pentland, B.T. 2003. Reconceptualizing organizational routines as a source of flexibility and change. *Administrative Science Quarterly* 48(1):94–118.

Finneran, E., P. Crosson, P. O'Kiely, L. Shalloo, D. Forristal, and M. Wallace. 2010. Simulation modelling of the cost of producing and utilising feeds for ruminants on Irish farms. *Journal of Farm Management* 14(2): 95–116.

Folke, C., S. Carpenter, T. Elmqvist, L. Gunderson, C. Holling, B. Walker. 2002. Resilience and sustainable development: building adaptive capacity in a world of transformations. *AMBIO: A Journal of the Human Environment* 31(5):437–440. doi: http://dx.doi.org/10.1579/0044-7447-31.5.437.

Ghosh, Malay, and J. N. K. Rao. 1994. Small area estimation: an appraisal. *Statistical Science*: 55–76.

Gloy, B.A. and E. LaDue. 2003. Financial Management Practices and Farm Profitability. *Agricultural Finance Review* 63(2):157–174.

Gordon, G. 1969. *System Simulation*. Englewood Cliffs, New Jersey: Prentice-Hall.

Green, S. and C. O'Donoghue. 2013. Assessing the Geographic Representativity of Farm Accountancy Data. *International Journal of Geo-Information*. 2 (1):50-66.

Griffith, V. 2014. PastureBase Ireland-the measurement of grass dry matter production on grassland farms. The future of European grasslands. Proceedings of the 25th General Meeting of the European Grassland Federation. 7–11 September. Aberystwyth, Wales. IBERS, Aberystwyth University.

Ham, Yoo-Geun, Michele M. Rienecker, Max J. Suarez, Yury Vikhliaev, Bin Zhao, Jelena Marshak, Guillaume Vernieres, and Siegfried D. Schubert. 2014. Decadal Prediction Skill in the GEOS-5 Forecast System. *Climate Dynamics* 42 (1-2): 1–20. doi:10.1007/s00382-013- 1858-x.

Hamrud, Mats, Bonavita Massimo and Isaksen, Lars. 2014. EnKF and Hybrid Gain Ensemble Data Assimilation. ECMWF Technical Memo 733. http://www.ecmwf.int/sites/default/files/elibrary/2014/9766-enkf-and-hybrid-gain-ensemble-data-assimilation.pdf.

Hanrahan, K., Hennessy, T., Kinsella, A. and Moran, B. 2014 Teagasc National Farm Survey Results 2013. Teagasc: Athenry. http://www.teagasc.ie/publications/2014/3179/NFS_2013_final.pdf.

Hennessy, T. and K. Heanue. 2012. Quantifying the effect of discussion group membership on technology adoption and farm profit on dairy farms. *The Journal of Agricultural Education and Extension* 18(1): 41–54.

Hennessy, T., Buckley, C., Dillon, E., Donnellan, T., Hanrahan, K., Moran, B. and Ryan, M. 2013. Measuring Farm Level Sustainability with the Teagasc National Farm Survey. Teagasc, Athenry, Co. Galway. http://www.teagasc.ie/publications/2013/3042/SustainabilityReport.pdf.

Hermes, K, and M. Poulsen. 2012. A review of current methods to generate synthetic spatial microdata using reweighting and future directions. *Computers, Environment and Urban Systems* 36(4): 281–290.

Hurtado-Uria, C., D. Hennessy, L.Shalloo, D. O'Connor, and L. Delaby. 2013. Relationships between meteorological data and grass growth over time in the south of Ireland. *Irish Geography* 46(3): 175–201.

Hynes, S., K. Morrissey, C. O'Donoghue and G. Clarke. 2009. Building a Static Farm Level Spatial Microsimulation Model for Rural Development and Agricultural Policy Analysis in Ireland. *International Journal of Agricultural Resources, Governance and Ecology* 8(3): 282–299.

Iversen, T., A. Deckmyn, C. Santos, K. Sattler, J.B. Bremnes, H. Feddersen, and I. Frogner, 2011. Evaluation of 'GLAMEPS'-a Proposed Multimodel EPS for Short Range Forecasting. Special Issue on probabilistic short-range weather forecasting. Tellus A 63(3): 513–30. doi:10.1111/j.1600-0870.2010.00507.x.

Janssen, S., and M. K. van Ittersum. 2007. Assessing farm innovations and responses to policies: a review of bio-economic farm models. *Agricultural Systems* 94(3):622–636.

Janssen, S., K. Louhichi, A. Kanellopoulos, P. Zander, G. Flichman, H. Hengsdijk, and M. K. van Ittersum, 2010. A generic bio-economic farm model for environmental and economic assessment of agricultural systems. *Environmental Management* 46(6): 862–877.

Kächele, H. and S. Dabbert. 2002. An economic approach for a better understanding of conflicts between farmers and nature conservationists-an application of the decision support system MODAM to the Lower Odra Valley National Park. *Agricultural Systems*74:241–255.

Kaine G.W. and P.R. Tozer. 2005. Stability, resilience and sustainability in pasture-based grazing systems. *Agricultural Systems* 83: 27–48.

Le Moigne, P., A. Boone, J. C. Calvet, B. Decharme, S. Faroux, and A. L. Gibelin. 2009. SURFEX scientific documentation. Note de centre (CNRM/GMME), Météo-France, Toulouse, France.

Lewis, P., J. Gómez-Dans, T. Kaminski, J. Settle, T. Quaife, N. Gobron, J. Styles, and M. Berger. 2012. An Earth Observation Land Data Assimilation System (EO-LDAS). *Remote Sensing of Environment* 120: 219–35. doi:10.1016/j.rse.2011.12.027.

Li, J. and C. O’Donoghue. 2014. Evaluating alignment methods in dynamic microsimulation models, *Journal of Artificial Societies and Social Simulation* 17(1):15. doi: 10.18564/jasss.2334.

Loughrey, J., Thorne, F., Kinsella, A., Hennessy, T., O’Donoghue, C. and Vollenweider, X. 2015. Market risk management and the demand for forward contracts among Irish dairy farmers. *International Journal of Agricultural Management* 4(4): 173–180(8). doi: http://dx.doi.org/10.5836/ijam/2015-04-173.

Louhichi, K., S. Janssen, H. Li, N. Borkowski, A. Kanellopoulos, M.K.Van Ittersum, G. Flichman, H. Hengsdijk, P. Zander, M. Blanco, G. Stokstad, I. Athanasiadis, A.E. Rizzoli, D. 2010a. A generic farming system simulator. In *Environmental and agricultural modelling integrated approaches for policy impact assessment,* edited by F. Brouwer, and M.K. Van Ittersum. 109–132. New York: Springer Academic.

Louhichi, K., A. Kanellopoulos, S. Janssen, G. Flichman, M. Blanco, H. Hengsdijk, T. Heckelei, P. Berentsen, A. O. Lansink, M. V. Ittersum 2010b. FSSIM, a bio-economic farm model for simulating the response of EU farming systems to agricultural and environmental policies. *Agricultural Systems* 103(8):585–597.

Macken-Walsh, A., K. Connolly, M. Gibson, K.Heanue, D. McCarthy, C. O'Donoghue, and C.Watson. 2015. Teagasc's eProfit Monitor: rationale, farmer uptake, and prospects, Teagasc.

Mahouf, Jean-Francois, and Gianpaolo Balsamo. 2015. Recent Developments in Land Data Assimilation for Numerical Weather Prediction. https://www.researchgate.net/publication/228882180_Recent_developments_in_land_data_assimilation_for_numerical_weather_prediction.

Mishra, A. K., H. S. El-Osta, and J. D. Johnson. 1999. Factors Contributing to Earnings Success of Cash Grain Farms. *Journal of Agricultural and Applied Economics* 31(3): 623–637.

Morrow, L., T. Kelly and T. Kirley. 2004. ICT-its potential as a channel for enhanced extension services. In Proceedings of 20th Annual Conference of Association of International Agricultural Extension and Education, Dublin, Ireland. May 2004.

Nelson, R. R. and S. G. Winter. 1982. *An Evolutionary Theory of Economic Change.* Cambridge, MA: The Belknap Press of Harvard University Press.

O'Donoghue, C., D. Ballas, G. Clarke, S. Hynes, and K. Morrissey. 2013. *Spatial Microsimulation for Rural Policy Analysis.* Springer-Verlag GmbH. ISBN 978-3-642-30026-4.

O'Donoghue, C., J. Lennon, K. Morrissey. 2014. Survey of Spatial Microsimulation Modelling. *International Journal of Microsimulation.* https://ideas.repec.org/a/ijm/journl/v7y2014i1p26-75.html.

Oliver, D. M., R. D. Fish, M. Winter, C.J. Hodgson, A.L. Heathwaite and D.R. Chadwick. 2012. Valuing local knowledge as a source of expert data: farmer engagement and the design of decision support systems. *Environmental Modelling & Software* 36:76–85. doi:10.1016/j.envsoft.2011.09.013.

Pannell, D. J. 1999. On the estimation of on-farm benefits of agricultural research. *Agricultural Systems* 61(2): 123–134.

Pentland, B.T. and M. S. Feldman. 2005. Organizational routines as a unit of analysis. *Industrial and Corporate Change* 14(5):793–815. doi:10.1093/icc/dth070.

Peters, Martin B., Enda O'Brien, Alastair McKinstry, and Adam Ralph. 2013. Wind Forecasting Using HARMONIE with Bayes Model Averaging for Fine-Tuning. *Energy Procedia* 40: 95–101. doi:10.1016/j.egypro.2013.08.012.

Seity, Y., P. Brousseau, S. Malardel, G. Hello, P. Bénard, F. Bouttier, C. Lac, and V. Masson. 2011. The AROME-France convective-scale operational model. *Monthly Weather Review* 139(3): 976–991.

Shadbolt, N., F. Olubode-Awosola, B. Rutsito, 2013. Resilience to 'bounce without breaking' in New Zealand Dairy Farm Businesses. 19th International Farm Management Congress, SGGW, Warsaw, Poland. ISBN 978-92-990062-1-4. http://ifmaonline.org/contents/pr-resilience-to-bounce-without-breaking-in-new-zealand-dairy-farm-businesses/.

Shalloo, L., P. Dillon, M. Rath, and M. Wallace. 2004. Description and validation of the Moorepark dairy system model. *Journal of Dairy Science* 87(6):1945–1959.

Shrestha, S., M. Abdalla, T. Hennessy, D. Forristal and M. Jones. 2014. Irish farms under climate change. Is there a regional variation on farm responses? *The Journal of Agricultural Science* 153(3): 385–398. doi:10.1017/S0021859614000331.

Shrestha, S., T. Hennessy and S. Hynes. 2007. The effect of decoupling on farming in Ireland: A regional analysis. *Irish Journal of Agricultural & Food Research* 46(1):1–13. http://www.jstor.org/stable/i25564549.

Simó, I., R. Corstanje, R. Fealy, S. Hallett, J. Hannam, N. Holden, and R. E. Creamer. 2014. Mapping soils in Ireland. In *GlobalSoilMap: Basis of the global spatial soil information system*, edited by Dominique Arrouays, Neil McKenzie, Jon Hempel, Anne Richer de Forges, and Alex B. McBratney. 167. London: Taylor Francis Group.

McEntee, P., R. Belford, R.Mandel, J. Harper and M. Trotter. 2013. Sub-paddock scale spatial variability between the pasture and cropping phases of mixed farming systems in Australia. In *Precision Agriculture 2013,* edited by John V. Stafford. 389–394. Wageningen Academic Publishers.

Tanton, R., Yogi Vidyattama, Binod Nepal and Justine McNamara, 2011. Small Area Estimation Using A Reweighting Algorithm. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 174(4):931–951.

Tess, M. W., and B.W. Kolstad. 2000. Simulation of cow-calf production systems in a range environment: II. Model evaluation. *Journal of Animal Science 78*(5): 1170–1180.

Teagasc 2015. eProfit Monitor Analysis. Teagasc, Oak Park, Carlow.

Van der Plas, E. V., B. Wichers Schreur, and K. Kok. 2012 A quantitative evaluation of the high resolution HARMONIE model for critical weather phenomena. *Advances in Science and Research* 8(1): 149–155.

Van Ittersum, M. K., M. MacLeod, S. Mayne, N. McRoberts, J. Oldham, A. Renwick and A. Wreford. 2009. Integration across disciplines: the lessons learnt from the integrated project SEAMLESS. *Aspects of Applied Biology* (93):55–60.

Villalba, D., I. Casasus, A. Sanz, A. Bernues, J. Estany, and R. Revilla. 2006. Stochastic simulation of mountain beef cattle systems. *Agricultural Systems* 89(2): 414–434.

Wallace, M. T. and J. E. Moss. 2002. Farmer Decision-Making with Conflicting Goals: A Recursive Strategic Programming Analysis. *Journal of Agricultural Economics* 53(1): 82–100.

Williamson, Paul, Mark Birkin, and Phil H. Rees. 1998. The estimation of population microdata by using data from small area statistics and samples of anonymised records. *Environment and Planning A* 30(5): 785–816.

Zander, P. 2001. Interdisciplinary modelling of agricultural land use: MODAM - Multi Objective Decision Support Tool for Agroecosystem Management. In *Multidisciplinary Approaches to Soil Conservation Strategies: Proceedings; International Symposium ESSC, DBG, ZALF*, edited by K. Helming. 155–160. Centre for Agricultural Landscape Research, Müncheberg.

Zhou, Yuhua, Dennis McLaughlin, and Dara Entekhabi. 2006. Assessing the Performance of the Ensemble Kalman Filter for Land Surface Data Assimilation. *Monthly Weather Review* 134 (8): 2128–42. doi:10.1175/MWR3153.1.

International Food and Agribusiness Management Review
Special Issue - Volume 19 Issue A, 2016

Big Data's Potential to Improve Food Supply Chain Environmental Sustainability and Food Safety

Mary Clare Ahearn[ⓘa], Walt Armbruster[b], and Robert Young[c]

[a] *Consultant and Senior Economist, Retired, U.S. Department of Agriculture, 2413 Mumford Dr. Silver Spring, MD 20906, USA*

[b] *Farm Foundation, Retired, 1709 Darien Club Drive, Darien, IL 60561, USA*

[c] *Chief Economist, American Farm Bureau Federation, 600 Maryland Ave., SW, Suite 1000W, Washington DC 20024, USA*

Abstract

Big data is emerging as an important information technology to guide decisions within agri-food supply chains. Big data can be used potentially to differentiate and identify final products based on underlying farm production attributes demanded by consumers in the supply chain. This paper considers the challenges faced by the supply chain in responding to consumer demands and adoption of big data technologies in agricultural production through closer evaluation of two examples—one of which considers the use of a sustainability metric and the other considers the potential to increase food safety. We conclude with some comments about likely future issues and implications of the potential widespread adoption of big data.

Keywords: big data, supply chain, sustainability, food safety

[ⓘ]Corresponding author: Tel: + 1 301.933.2112
Email: M. C. Ahearn: mahearn54@aol.com
W. Armbruster: walt@farmfoundation.org
R. Young: boby@fb.org

Introduction

Recently, in *A Framework for Assessing Effects of the Food System*, the National Research Council (2015) offered a general framework for assessing the entire food system. The framework moves beyond environmental sustainability, and considers the health, social and economic domains. However, this expansive framework is a long way from quantifying production practices that can be certified, as required to realize a consumer-driven supply chain. A scientific literature does exist on the development and validation of indicators of individual aspects of sustainability in selected farming systems, but these are generalized and not site-specific (e.g., see Bell and Morse 2008). Long term efforts have shown economic and environmental benefits of specific production practices facilitated by precision agriculture techniques. Similarly, the first point to keep pathogen contamination out of the food supply chain is in the production fields, where precision agriculture can facilitate monitoring and limit foodborne pathogens—especially important with fresh produce which receives minimal processing beyond the farm gate. However, on-farm practices are very site specific, and translating research into generalizations about safe and sustainable practices is problematic. Further, the remaining nodes in the food supply chain must be addressed in any system designed to increase the level of food safety.

Big data offers a technological breakthrough that may provide a means for translating "good" practices into generalizations that consumers can trust and be willing to pay for and, at the other end of the supply chain, firms could use for monitoring and evaluation of alternative solutions to providing sustainably produced and safe food. Even though consumers would be the ultimate beneficiaries, it is the intermediaries in the food supply chain who must identify and develop or adapt existing data sources needed to operationalize best practices. Developing means of successfully capturing the big data being created in the production process and analyzing it to create valuable tools and metrics for use by managers in production and supply chain firms requires new analytics adapted to the particular issues involved. In spite of the measurement challenges, there is growing interest on the part of major players in the supply chain to meet consumer expectations.[1] The challenges faced by food supply chain managers in responding to consumer demands are illustrated here through two examples: one considers the use of a sustainability metric and the other considers the potential to increase food safety. We conclude with some comments about likely future issues and responses of agri-food supply chain managers.

Evolution of Consumer Demand for Food Products

The business of farmers and food supply chains has traditionally been to provide consumers with food products that meet their marketplace demands for quality and affordability. Indeed, recent surveys indicate that consumers are most concerned with affordability, nutrition and food safety (Glassman 2015). There is also a long history, particularly in developed countries, of governments playing a role in encouraging (through incentives) and requiring (through regulations) agri-food supply chains to meet certain standards in their production processes. One justification for government involvement in the marketplace is the public good aspects associated with the production of the final products demanded by consumers. For example, the government has a shared responsibility to minimize foodborne

[1] For example, food chains make efforts to be listed as part of the Global 100—of the world's most sustainable corporations—announced annually at the World Economic Forum. Among the 100, four food companies were listed most recently: General Mills, Unilever, Coca-Cola, and Campbell's Soup.

disease outbreaks and environmental degradation. In both cases individual firms, particularly in the short run, have limited financial incentive to pursue them.

As evidenced by the contemporary foods movement (e.g., Pollan 2006), fueled at least in part by information technology (Streeter, Sonka, Hudson 1991; Poppe et al. 2013), consumers are paying closer attention to the products and processes of the food supply chains. For example, consumers are educating themselves about the dangers of foodborne diseases, as evidenced by an increase in internet searches following government reports of outbreaks (Kuchler 2015). In addition, consumers are widening their interests in the public good aspects of the food production systems, including environmental impacts of alternative production processes. Perhaps in response to the lack of government involvement in the food system or to preempt government involvement, food supply chains are increasingly becoming engaged in the provision of what heretofore have been considered public goods. At least in some part, firms are also pursuing the effort for marketing purposes (Elder and Dauvergne 2015).

The attributes that define a food product have recently expanded. For example, consumers are now offered a variety of egg products differentiated by the labeling of the on-farm production processes, such as produced by cage-free and free-range hens. However, this has also raised a concern that, especially as industry concentration increases, consumer choice may be restricted as a result of corporate decisions to limit the offering of products that respond to the demands of only a subset of consumers. Recent corporate decisions on the part of retail and fast food chains regarding cage-free eggs is an example. A recent article by Saitone, Sexton, and Sumner (2015) considers just such a case when a market response to consumer interests in food production processes can have the effect of limiting consumer choice and increasing costs. This occurs when players in the supply chain offer only selected food items produced using specified processes instead of offering a selection of products with alternative bundles of characteristics. In their study of antibiotic-free pork production, using simulations, they conclude that the increase in production costs led to significant reductions of both consumer and producer surplus. It is worth noting that the study focused only on private returns and not the public good aspects of the development of increased antibiotic resistance due to the use of antibiotics in pork production.[2]

The Corporate Awakening

As a more connected and informed consumer base has evolved, so too have agri-food supply chains in response to that evolution. Food supply chains can respond in a variety of ways to a more interested consumer population. Developing a positive reputation among the consumer base—so-called self-regulating—is important as a defensive strategy to possibly avoid costly government regulation or agency costs, i.e., costly lawsuits. It is recognized in the business management literature that firms are interested in demonstrating to their customers, through claims of Corporate Social Responsibility (CSR) that they care about the social welfare and environmental impacts of their businesses (Stephen 2004). In short, CSR involves spending more doing business than is required by law and regulation to accomplish a public goal (McWilliams and Siegel 2001; Paul and Siegel 2006).

Corporate managers generally have control over their firm's discretionary spending to exhibit CSR and the actions take many forms, from supporting local social causes near their headquarters to engaging in social or environmental activities generally related to their

[2] For a review of the evidence on these effects, see Teillent and Laxminarayan 2015.

industries. Firms may convey their responsiveness through general declarations of corporate social responsibility (CSR) which are not unlike the more traditional philanthropic donations (Hay, Stavens, and Vieter 2005). Major food companies have stated their intention to be more socially responsive by how they manage their businesses. For example, in the buildup to the United Nations Climate Change Conference meeting in Paris, November 2015, major agri-food companies were fully engaged in declaring their support of reducing greenhouse gases. This includes Cargill, PepsiCo, Wal-Mart, General Mills, Hershey, Kellogg's, Mars, McDonald's, Monsanto, ethanol firms Abengoa and Poet, and Campos Brothers Farms (Basher 2015). Another example of CSR from the food supply chain is Wal-Mart's goal to increase local foods sales—in some areas sourced from limited resource farmers—even though it is not established that local foods are superior nutritionally or lead to less environmental degradation.

There is not strong scientific evidence about the relationship between CSR claims and sustainability accomplishments and, in fact, there is some evidence that CSR claims are not improving food security and sustainability in developing countries (Elder and Dauvergne 2015). Similarly, there is little empirical evidence on how CSR claims affect consumer demand for food items. Moreover, given that agricultural supply chains are generally global in nature, the response of consumers to claims of CSR are expected to vary significantly.[3] While actions of CSR may buy the food industry good will, it may not be sufficient to meet consumer demands focused on the food products they purchase for attributes relating to food safety and claims regarding the underlying sustainable farm production processes. An entire industry has grown up regarding these claims and audits associated with verifying those claims. It is the potential availability of large data sets that may allow these claims to be made and verified, particularly in a commodity industry.

Reassurances to meet these more specific demands are only possible when there is an established scientific basis and a system designed to capture production information from the farm to the food product. The collection of location specific, auditable information represented by big data is poised to play a major role in that development. However, a potential danger to innovation in the supply chain could result from food processers pushing to become more consumer-driven. This would be the case if it leads to an excess of centrally-defined production paths to accomplish goals, which thereby crowd out historical farm-level innovation.

Big Data in Context

The term *big data* is used in a variety of contexts, inside and outside of agriculture, and is very broadly described. No satisfactory general definition exists. As noted in a recent National Research Council report on the topic, no satisfactory definition can be provided until there are general laws established that are scale neutral in their applicability (Committee on the Analysis of Massive Data, 2013). Descriptions of big data in the agricultural context generally emphasize extremely large data sets (generally built by integrating multiple sources of related data), analyzed with state-of-the-art computing power to reveal patterns, trends, and associations of value for a variety of decision making purposes. Our emphasis in this paper is

[3] For example, in an analysis of the impact of CSR claims on wine sales internationally, Muellor-Loose and Remaud (2013) found significantly different impacts across 5 developed countries. They also found that claims of CSR were valued less by consumers than the organic label.

in the use of large amounts of data integrated by those in the agri-food supply chain to provide consumers products with desirable attributes.

On-farm data collection related to detailed production practices, input use, disease outbreaks, food safety concerns, and yields has been occurring through the use of increasingly sophisticated machinery and equipment, often termed precision agriculture. Additional crops will benefit from these technologies as researchers work with industry to develop systems for particular crops. For example, the grape industry expects to benefit from precision vineyard management being developed under a current U. S. Department of Agriculture (USDA) National Institute of Food and Agriculture (NIFA) Specialty Crop Research Initiative grant. By collecting and analyzing thousands of images per minute, farmers will be able to hone their management practices to improve quality and quantity of their crops (Enos 2015). This should facilitate monitoring for pathogens or field conditions particularly susceptible to creating potential food safety risks, and taking remedial action or even removing suspect product from the supply chain at the farm level. Operators using this equipment and the sensor-based information generated have the potential to tailor seed variety, crop nutrients and other production practices down to a resolution of a few feet. This may help to improve profitability as well as reduce the operator's environmental footprint. It may also provide the input needed to evaluate sustainability metrics many companies are turning to in order to provide objective, measurable evidence of improvements in the environmental performance of a company's supply source.

The unique contribution of big data is to combine the vast amount of data from public investments, such as weather and climate predictions of major models, with aggregations of on-farm input and output relationships from relevant regions to develop alternative management strategies for desired outcomes. Industries associated with agriculture are finding creative ways to add value to these aggregated data, including selling management services to producers and monitoring the practices of their upstream producers to be used in the marketing of their products with specified attributes in their supply chain. However, there are still significant legal issues to be resolved because the current laws addressing intellectual property do not provide a clear interpretation for agricultural data as it relates to trademark, patent, or copyright law (Ferrell 2015). Farmers also have expressed concern about giving up the property rights to their own data (Thatcher 2015). In response to this concern, one tool was developed by a coalition of farm, commodity, and agricultural technology providers—The Ag Data Transparency Evaluator—to help producers understand where their data is going and who has access and control over it. Adequately addressing the legal issues, however, may require congressional action to revise the Uniform Trade Secret Act to encompass the uniqueness of agricultural data. Some farmers have pursued concerted action that can make their individual data valuable to them in assuring sustainable production practices and monitoring food safety practices at the local level. One strategy is banding together in cooperatives to aggregate and analyze their data, working with universities to develop analytic platforms. They can then protect the privacy of their individual data while sharing the added value within the cooperative membership. Such private sector actions, collaboration between farmers and providers of digital information systems, and/or public policy will need to sort out the privacy and legal issues involved.

The most recent development to help farmers manage their data and capture the value to them in documenting characteristics of their operation—which could include sustainability and food safety related metrics—is the formation of an Agricultural Data Coalition (ADC 2016). It is "focused on designing, creating and managing a central repository where farmers can store their information and oversee how it is accessed”. This will potentially allow farmers to maximize the value of their data by using it to accomplish their own goals. At the same time, it addresses some of the ethical issues of control and use of big data at the farm level end of the food supply chain.

Two Examples of Consumer-Driven Supply Chain Responses

While the primary motivation for a farmer to adopt the use of precision agriculture or big data is profitability, agri-food supply chains are poised to capture the information from production systems to make attribution claims of value to their customers. We consider two distinct industry responses: The first is industry responding to consumer demands for more environmentally sustainable grain production and the second involves industry responding to both consumer demands and federal regulation regarding food safety practices. It is interesting to note the differences between the two responses in the level of scientific knowledge of the outcome to be avoided, i.e., environmental degradation and negative health outcomes and even death due to pathogen risks in food. For food safety, when a significant outbreak resulting in deaths or illnesses occurs, the Centers for Disease Control makes an effort to identify the source of the outbreak and the federal government and industry respond. However, the system is in the process of changing since the final rules under the 2011 Food and Drug Administration's (FDA) Food Safety Modernization Act (FSMA) became effective in January 2016. The FSMA turns the primary focus to prevention rather than reacting to correct foodborne illness incidents and deaths and reduce costly recalls (FDA 2011). In contrast, it is more difficult to identify and measure an acceptable level of environmental degradation. Moreover, in contrast to a response to food safety concerns, solutions to environmental issues are less clear because it is difficult to identify simple trade-offs among on-farm production practices. The environmental situation is a classic example of what has been characterized as a "wicked problem" in agriculture (Batie 2008).[4]

The Maturation of the Sustainability Concept

The general and evolving concept of sustainability does not lend itself to simple measurement. The 1987 United Nations Report of the World Commission on Environment and Development, gave a definition of sustainability as “… implies meeting the needs of the present without compromising the ability of future generations to meet their own needs” (United Nations 1987). In the report, three interrelated features were noted: economic viability, social equity, and environmental protection. The National Research Council in its 2010 report, *Toward Sustainable Agricultural Systems in the 21st Century*, proffered that sustainability meets four goals: (1) To satisfy human food, feed, and fiber needs, and contribute to biofuel needs. (2) To enhance environmental quality and the resource base. (3) To sustain the economic viability of agriculture. (4) To enhance the quality of life for farmers, farm workers, and society as a whole. Sustainability is the greatest, the report also hypothesized, when a generalized set of production processes has the greatest overlap of these goals. Implementation of the environmental goals of sustainability is particularly challenging for agriculture because of the multitude of factors that are relevant in measuring the impact of

[4] An early literature on Post-Normal Sustainability Technologies is exploring how to frame wicked problems in public policy (Funtowitz and Ravetz 1990).

production practices. Improvement in one dimension of environmental performance, may come with environmental degradation in another (Aigner, Hopkins, and Johansson 2003).[5]

The quest for sustainability has become a focus for consumers. It has also become a driver of innovation in the agri-food supply chain. In a recent Harvard Business Review article, sustainability is seen as driving innovation to meet new regulatory standards and create value chains in cooperation with downstream and upstream partners (Nidumolu, Prahalad, and Rangaswami 2009). As the concept of sustainability has matured, the agri-food supply chain is at the initial stages of utilizing the tools of big data to more efficiently evaluate farmers' actions in reducing their environmental footprint. These evaluation systems can take considerable time. For example, the "Field to Market" sustainability calculator takes as much as thirty minutes to populate the model for one field. That information provides measures of soil erosion, energy and water use and crop nutrient efficiency among others. To as great an extent possible, the calculator pre-populates many of the required data inputs such as soil type or field slope by linking to federal data sources, but the production practice information still requires the farmer to spend input time. To provide that information for the entire farm may well take hours. Linking information directly from the farm's machinery compliment will one day hopefully negate the need for this kind of time commitment as well as providing for a much more detailed look at the farm's environmental performance. It may also allow the farmer to self-develop best management practices for the operation using the farm data itself as a set of replicated experiments. From a CSR perspective, the company purchasing the farm's product would have access to detailed environmental measures as well as data backing up any continuous improvement claims.

While sustainability purports to embrace goals of economic profitability, natural resource conservation, and quality of life, in practice, quality of life goals are often ignored in developed countries. In developing countries, agricultural development and rural development are more closely linked since rural livelihoods and agricultural profitability are often one and the same, but they often rank low on quality of life indicators. The balancing of the triple bottom-line of sustainability—economic, environmental, and social—is very much a work-in-progress for agri-food supply chains which are global in nature. For example, Hidayat, Glasbergen and Offermans (2015) analyze the implications of sustainability certifications developed to meet the environmental goals of one regional market but where the production occurs in another region that experiences significant social impacts to rural livelihoods.

Sustainable Label in Grain Production

Translating sustainability into the marketplace for agri-food products is driving companies as well as farmers and ranchers to re-evaluate what have been traditionally viewed as good farming practices. It is also bringing into question policy approaches to environmental performance improvement efforts that have been in place for decades. One of the major contributions of big data will be to help companies and agriculture measure the sector's

[5] The International Organization for Standardization (ISO) is the world's largest developer of standards, currently with 162 member countries and standards dedicated to food production, sustainable development, water, and other areas of relevance to agricultural production. The Sustainable Agriculture Initiative (SAI) is an organization established by the food industry which supports sustainable agricultural practices. The challenges in implementing these types of certification of standards have been documented (e.g., by UNEP 2000) and sometimes criticized by consumer and environmental groups (e.g., Friends of the Earth).

performance, while also granting individual companies the data paths they feel are necessary to make sustainability claims.

The interface between CSR claims and commodity agriculture continues to create challenges for all involved. Companies want to be able to make claims regarding their individual activities toward their CSR goals. For identity preserved product, these claims are fairly easy to make and verify. With Price Lookup Code (PLU) stickers on individual pieces of fruit or vegetables along with coding on the box itself it is possible to trace a head of lettuce back to the row of the field from which it was harvested and in many cases even down to who was working on the crew that particular day. For products with PLU codes the entire claim and verification process is straightforward. Identity preservation claims for what are usually viewed as bulk commodities can be relatively expensive.

One of the most widely used class of products are organic grains and oilseeds. At the end of September 2015 the Agricultural Marketing Service of the USDA reported the national average price of organic corn at $10.74/bushel while at the same time corn was selling for $3.66 per bushel in Chicago. Simultaneously, organic soybean prices were reportedly at $21.81 per bushel while conventional beans were at $8.62 per bushel (USDA 2015). A twenty-five pound bag of tofu quality soybeans was available through Amazon for $68.88 plus shipping in early October 2015. Again, the organic designation has a large enough market share to have its own distribution, price discovery and marketing channels. For an individual company trying to establish their own supply chain for a given set of production practice requirements, the costs are quite large at nearly any kind of scale.

There are companies, however, with sufficient market share, to make demands that may require the entire chain to consider alterations in production practices. A 2013 Forbes article pegged Wal-Mart with a 25% share of national grocery sales (Leeb 2013). Wal-Mart also has a well-developed sustainability effort, requiring their suppliers to devote considerable effort to document their work toward reducing greenhouse gas emissions, water usage and other environmental measures (Wal-Mart 2015). Other consumer-facing companies such as Unilever have well established sustainability programs and are working on pilot programs to source sustainable soy in the United States, using Unilever’s sustainable Agricultural Code (Unilever 2015). But the pilot discussed in the Unilever case covered only 160,000 acres in 2014 according to this same website and 83 million acres were planted to soybeans in the United States in 2014. Unilever has stated a goal of having a million acres enrolled in a sustainability calculator, "Field to Market," by 2017. Data collection on a farm-by-farm basis through surveys is not without cost, which is exactly why many are looking to big data and automated data collection systems as the mechanism whereby the sector and the companies will be able to make sustainability claims.

Sustainability claims for many products require a producer to go through an extensive checklist, reporting on everything from nutrient use to labor practices. The Field to Market program discussed by Unilever—and looked to by many in the commodity crop space—utilizes a set of metrics that currently require significant input from the farmer to complete. These are then benchmarked against other data from the local area or a group of other selected producers who are usually participating in the same pilot program. Field to Market currently has more than eighty groups (grower organizations; agribusinesses; food, fiber, restaurant and retail companies; conservation groups; and universities) as partners (Field to Market 2015).

There is an effort underway to convert these same metrics so that they will interact directly with data management systems consistent with those maintained for giving cropping prescriptions. This big data approach means the data accuracy would reflect that collected straight from the piece of operating equipment and would significantly reduce the need to audit the data collected. If, or when, this can be taken through to execution it would allow the sector to make the same kind of sustainability claims over tens of millions of acres quickly as opposed to the few hundred thousand acres currently enrolled. However, computer scientists are only beginning to address the technical infrastructure challenges presented by big data analysis which will require hardware and software advances in both parallel and distributed processing systems. While additional observations and new data sets can improve the ability to address a problem and/or expand applications, their processing infrastructure also raises challenges due to heterogeneity in representations, data quality, and openness (Committee on the Analysis of Massive Data 2013, Chapter 4). Concerns about data ownership and confidentiality, in particular, have surfaced early in the development of big data technologies for agriculture. Computer technologies are needed which protect the privacy of confidential data from leakage and malicious harvesting, while fusing various data sets, through hardware parallelism.

Enhancing Food Safety in the Supply Chain

Food safety throughout the supply chain is an ongoing concern. Though much effort goes into assuring safe practices in food production and handling, a large number of foodborne illnesses and deaths are experienced annually within the United States. One estimate of the cost of foodborne illnesses is $56–$93 billion annually, based upon immediate treatment cost and lost income of individuals but not accounting for potentially significant costs related to long-term health impacts (Scharff 2015). Contamination may occur at any point in the food supply chain starting at the farm level, thus creating challenges in identifying the cause of foodborne illness outbreaks. Since there are many more instances of foodborne illness than are identified and traced to a source, the private market incentives to optimize food safety are somewhat weak across the entire food supply chain. A lack of information available to consumers, industry, and policy makers creates problems in preventing foodborne illness from pathogens, but there are several options to provide better information. These include more financial support of databases and research by federal agencies; a farm-to-table database to trace pathogens to specific farms, food companies, and products; creation of a national product liability database documenting court cases including out-of-court settlements for foodborne illness cases; and establishing a Cabinet level or independent consumer protection agency to foster collection of data on pathogens in the US food supply (Roberts 2013). However, there are significant incentives for preventing foodborne illness, especially for major branded retailers and food service establishments where the value of their brand and franchise could be greatly diminished or destroyed by a serious foodborne illness incident. For example, sales in established, i.e., open a year or more, Chipotle Mexican Grill, Inc. restaurants fell 30% in December 2015 from the previous year. Chipotle stock market value fell to as low as $400 per share from over $750, a 53% decline in implied company market value, following a series of E. coli and norovirus food safety incidents in its popular restaurants, starting in October 2015 (Jargon and Newman 2016).

One means of protecting their brand is for a company to incorporate a strong risk management strategy as a critical part of its business plan (Brackett 2015). Denmark has used a collaborative approach between government and industry to eliminate salmonellosis in the poultry supply chain, which had become a serious problem. According to the vice president for food safety at Cargill Turkey and Cooked Meats, they already find it in their best interest to control what goes on at the farm and any new regulatory requirements are not likely to much change what they are already doing (Clapp 2015). Given the critical nature of food safety to a company's brand and ability to participate in the supply chain for such brands, most food companies undoubtedly have a somewhat similar philosophy and practice. But the voluntary nature of the US food safety system and frequently a lack of access to data regarding safety in the supply chain means that challenges remain.

As mentioned, the current voluntary food safety system is about to change under the 2011 FDA Food Safety Modernization Act (FSMA), which is the first update to US food safety laws since 1938. It gives FDA more authority to regulate fresh produce and animal feed, food imports and transportation, and provide oversight of third-party auditors. Sec. 206 of the Act provides the FDA with authority to mandate recalls of contaminated food if the responsible party does not cease distribution or recall a product (FDA 2011). The final rule on preventive controls for human food under FSMA establishes key requirements and compliance dates by which "covered facilities must establish and implement a safety system that includes an analysis of hazards and risk-based preventive controls", documented in a written safety plan. The rule provides flexibility in oversight and management of preventive controls. This must include "monitoring ... appropriate to the preventive control", corrective actions for "a minor, isolated problem that occurs during food production", and verification "that preventive controls are consistently implemented and effective." It clarifies that Primary Production Farms and Secondary Activities Farms are not subject to the preventive controls unless they handle produce covered by the Produce Safety Rule with which they are required to comply. It also "mandates that a manufacturing/processing facility must have a risk-based supply chain program for those raw material and other ingredients for which it has identified as a hazard requiring a supply-chain applied control" program. The final rule also updates and clarifies Current Good Manufacturing Practices (CGMPs) (FDA 2015).

One approach to increase incentives and accountability for improved food safety is to identify good manufacturing practices for food processors and handlers at various stages in the food supply chain. There are examples of industry efforts to establish such initiatives within their sphere of influence, relying on voluntary compliance with guidelines developed by industry organizations. An example is guidelines which provide recommended food safety practices for the fresh tomato supply chain which are intended to minimize microbial hazards associated with fresh and fresh-cut tomato products. The North American Tomato Trade Work Group (NATTWG) and the United Fresh Produce Association provided leadership for this effort involving a number of associations, agencies, companies and individuals with expertise in food safety practices. The guidelines are divided into eight primary modules starting with open field production through food service and retail. Each module addresses key considerations to control potential sources of pathogen contamination reasonably likely to occur in the absence of such control. These guidelines are intended to enhance safe growing, processing, distribution and handling of the commodity from field to consumer, supplementing existing food safety programs encompassing Good Manufacturing Practices (GMPs) and Hazard Analysis Critical Control Point (HACCP) programs (Gombas et al. 2008). A proliferation of sensor technologies to gather big data are available for use throughout the food supply chain from farm to field to consumer point of purchase. With appropriate analytics to provide needed information, the data can facilitate management application of the guidelines to enhance food safety. Industry organizations could further incentivize compliance through educational efforts within the industry and publicizing the existence of the guidelines to consumers, simultaneously monitoring compliance and recognizing those firms formally adopting the guidelines.

A public sector approach would be to specify a set of required actions for food processors and handlers, which would then be monitored continuously at the various nodes in the supply chain. The monitoring could be a direct government function as in the case of a number of U.S. Department of Agriculture (USDA) programs related to food safety, product quality grades, and standards of product identity. Alternatively, the USDA could certify third-party organizations to carry out the ongoing monitoring, as is currently done to assure that food sold as organic is indeed produced in compliance with the National Organic Standards Act requirements. Making public the names of firms violating food safety requirements could incentivize compliance (National Research Council 2011), or a system of fines could be put in place for violation of requirements. This kind of continuous monitoring—by either the private or public sector—would produce prodigious amounts of data from which innovative analytics would be required to sort out useful information, i.e., a big data solution. The advantage would be the possibility of heading off potentially calamitous foodborne illness events that could affect from only a few to thousands of consumers with various degrees of illness severity, avoiding long term health impacts, and mitigating significant private and public cost consequences.

Either of the above approaches would require prioritizing interventions to control pathogens. Risk assessment and cost benefit analysis can be used to evaluate pathogen interventions in the food supply chain and provide a basis for such prioritization. This could provide a basis for increasing information available to consumers as well as businesses throughout the supply chain. The most stringent approach would establish databases tying specific foodborne illnesses to particular food producers, products or companies (Roberts 2015).

As pointed out earlier, the initial place to keep pathogen contamination out of the food supply chain is in the production fields, particularly with fresh produce which receives minimal processing beyond that point. An effort to create a GIS-based online tool to identify specific points in a field where foodborne pathogens are prevalent can provide growers a risk-based strategic approach to focus food safety attention. Heat maps pinpointing relative levels of risk in each field would reduce the difficulty of successfully adopting this approach at the farm level to keep potential pathogens from entering the food supply chain (Wiedmann 2014). A Cornell university project funded by the Center for Produce Safety is working to create a modeling tool which is GIS-based to identify specific locations within fields where risk of pathogen contamination may be higher. The GIS tool will utilize unique characteristics of a farm including soil moisture and precipitation, two big data driven elements, to identify areas of the farm they should target to employ science-based strategies to mitigate risk of contamination. Another Center for Produce Safety project will develop an application for computer or cell phone use to allow producers to minimize the illness outbreaks from E. coli and salmonella contamination from irrigation water, a frequent source of foodborne illness in fresh produce. A model will help growers determine the need for increased risk-based sampling based on rainfall, irrigation methods and type of produce. Utilizing local weather information will make this approach usable in many different areas of the country to determine needed frequency of sampling following rainfall which is a significant influence on surface water quality (Rock 2014).

At the industry level, public and private sector networks to analyze the data tracing food safety outbreak causes exist within countries and in international contexts. The 2011 FDA

Food Safety Modernization Act created incentives for the US private sector to adopt HACCP practices which include properly cleaning processing plants and keeping foods properly refrigerated during transport. Low cost but accurate sensors for continuous monitoring allow companies to strengthen food safety.

Scanning equipment in plants, ubiquitous personal devices, shipment tracking, and retail monitoring of consumer purchases creates the big data with potential to enhance traceability throughout the supply chain (Armbruster and MacDonell 2014). The potential to trace specific lots of food from a particular node in the supply chain to a specific location on the retail shelf or food service establishment receiving food identified as potentially contaminated is within reach. This could greatly reduce the costs of recalls relative to a complete removal of all product produced by a firm during a given time period. Speeding up the recall and making it more focused could increase the chances of removing the food before consumption rather than notifying the public after much or at least some amount of the potentially contaminated food has been consumed.

The rapid growth in demand for animal-based foods and vegetables—both rather risky for food safety—in rapidly emerging economies is a particular concern for preventing foodborne illnesses. Further, intensification of agriculture to meet this growing demand in countries where governments systems strain to keep up with rapid growth is of particular concern. Given that food safety and prevention of foodborne illnesses are global public goods, international cooperation and investment to ensure safer foods will be needed both by international organizations and national governments (Grace and McDermott 2015). The FSMA contains more rigorous food safety requirements for US imports. The implementation and enforcement of those provisions will determine the extent to which it adequately protects consumers from unsafe food imports over time.

Key to utilizing big data to improve food safety is much faster pinpointing of foodborne illness outbreak causes and sources; relevant technologies are being developed throughout universities and the private sector. For example, a particularly promising technology is a machine using optical scanning, laser sensor technology developed by Purdue University. It is capable of pinpointing eight specific strains of salmonella as well as identifying a number of the other most important foodborne illness pathogens with an accuracy greater than 95%. Its big advantage is the ability to produce results within twenty-four hours, as opposed to the current industry-standard of seventy-two hours. It has the potential to provide an inexpensive, efficient preliminary screening tool, an appealing prospect for the food processing sector which should lead to rapid adoption of this technology once it is perfected for use (van Hoose 2015). The implementation of this optical scanning, laser sensor technology could be very valuable in preventing potential foodborne illness incidents through screening processing/ handling/transportation operations in the supply chain where monitoring would produce voluminous data points to allow pinpointing a problem. Appropriate analytics would be required to process the data for real time decisions, and allow connecting any problems with earlier points in the supply chain which may contribute to them. By preventing potential incidents of food borne illness, the consumer would be directly impacted by avoided health impacts and related costs. The firms implementing such a system would need to publicize the extent to which they were going to prevent food borne illness, since the process would not be apparent or observable by consumers.

Concluding Comments

Consumers increasingly want to purchase food produced with certain underlying farm production practices, while having confidence in the safety of the food. The food industry increasingly wants to provide consumers with reassurances about these characteristics, while maintaining the efficiencies for which the industry is known. Farmers, too, want to continue to earn their reputation as good stewards of the land producing high-quality, safe, and affordable food at profitable levels. However, there are a number of barriers to the implementation of these goals in the supply chain. This paper has addressed the potential for the use of big data to help overcome some of those barriers and improve the performance of supply chains in meeting consumer demands.

Returns from scientific investments, good practices, field-level precision agricultural techniques and, of course, the underlying technology boom in general, have allowed for the emergence of the possibility of using big data to improve the efficiencies of the agri-food supply chain. A barrier to marketing products as nutritious, safe, and sustainably produced is that the science is not clear. In addition, consumers are sometimes saddled with misinformation. While there will always be a lag from knowledge generated in the scientific lab to adoption in the supply chain, big data can help to reduce that diffusion time and strengthen confidence in the results. We highlighted two examples of emerging supply chain responses to consumer demands for attributes associated with production processes and the quality and safety of the final product. To date, science-based regulation has played a larger role in providing incentives to supply chains regarding food safety objectives, compared to environmental sustainability objectives. In contrast, the linking of environmental attributes to food products has largely been initiated within the supply chain in response to consumer demands, either through the certification of attributes which are priced into the final food product or as part of a company's Corporate Social Responsibility agenda. As governments around the globe begin to respond to climate change outcomes in the form of international agreements to reduce carbon emissions, the supply chains could see more incentives provided by government regulation. Although the two examples highlight different roles for government as a result of the underlying scientific challenges, big data is poised to play an even greater role in the efficiency of the supply chains.

There are often unintended consequences of any technology adoption. In the adoption of big data as a tool in organizing and managing agri-food supply chains, we have mentioned three. First, big data in concentrated global supply chains may lead to a loss of consumer options, the hallmark of thriving markets. This is especially true in the case of products with environmental sustainability attributes, since most consumers are less willing to incur the private risks associated with foodborne illnesses than they are the less certain risks associated with long-term environmental degradation. Secondly, if not managed appropriately, there is a danger that the traditional source of innovation from individual farm-level management strategies will be lost to the supply chain through over prescription of farm practices. Finally, the adoption of big data technology on the farm may not be scale neutral (or spatially-neutral). Part of the stated triple bottom-line of sustainability is the focus on social equity, oftentimes interpreted to mean a farming system without extreme production concentration. Adoption of big data technologies is likely to accelerate the concentration in production agriculture.

References

Agricultural Data Coalition. 2016. Agriculture Data: Putting Farmers in the Driver's Seat. http://agdatacoalition.org/. [accessed 3/13/2016].

Aigner, R., J. Hopkins, and R. Johansson. 2003. Beyond Compliance: Sustainable Business Practices and the Bottom Line. *American Journal of Agricultural Economics* 85 (5): 1126-1139.

Armbruster W. J. and M. M. MacDonell. 2014 Informatics to Support International Food Safety, Proceedings of the 28th EnvironInfo 2014 Conference, Oldenburg, Germany: September 10-12. http://enviroinfo.eu/sites/default/files/pdfs/vol8514/0127.pdf.

Basher, P. 2015. Food, ag firms pledge to combat climate change. Agri-Pulse Communications, Inc. October 19.

Batie, S. 2008. Wicked Problems and Applied Economics. *American Journal of Agricultural Economics* 90(5): 1176-1191.

Bell, S. and S. Morse. 2008. *Sustainability Indicators: Measuring the Immeasurable?* Second edition. London, UK: Earthscan.

Brackett, R. E. 2015. Food Safety Considerations for Nutrition Science. Academy eBriefings. *The New York Academy of Sciences*, Feb. 6. www.nyas.org/FoodSafety-eB

Clapp, S. 2015. Denmark embarrasses U.S. when it comes to Salmonella. *Food Chemical News* 23-April. https://www.agra-net.com/agra/food-chemical-news/food-safety/outbreaks-and-recalls/denmark-embarrasses-u.s.-when-it-comes-to-salmonella-476794.htm.

Committee on the Analysis of Massive Data; Committee on Applied and Theoretical Statistics; Board on Mathematical Sciences and Their Applications; Division on Engineering and Physical Sciences; National Research Council. 2013. *Frontiers in Massive Data Analysis.* Washington, D.C. National Academies Press.

Elder, S. and P. Dauvergne. 2015. Farming for Wal-Mart: The Politics of Corporate Control and Responsibility in the Global South. *Journal of Peasant Studies* 42(5):1029–1046.

Enos, L. 2015. National Grape & Wine Initiative Announces $6 Million Federal Research Grant to Transform U.S. Vineyard Management. PR Web, Oct. 8, Sacramento, California.

Ferrell, S. 2015. Big Data and Agriculture: Innovations and Implications. Testimony to the Committee on Agriculture, U.S. House of Representatives, October 29.

Field to Market. 2015. The Alliance for Sustainable Agriculture. About Us. https://www.fieldtomarket.org/.

Food and Drug Administration. 2011. Full text of the Food Safety Modernization Act (FSMA). http://www.fda.gov/Food/GuidanceRegulation/FSMA/ucm 2475 4 8.htm#SEC206.

Food and Drug Administration. 2015. Final Rule on Preventive Controls for Human Food. http://www.fda.gov/downloads/Food/GuidanceRegulation/FSMA/UCM461834.pdf

Funtowitz, S. and J. Ravetz. 1990. *Uncertainty and Quality in Science for Policy*. Kluwer Dortrecht.

Glassman, M. 2015. Hungry for Information: Polling Americans on Their Trust in the Food System. The Chicago Council on Global Affairs. October.

Gombas, D., et al. eds. 2008. *Commodity Specific Food Safety Guidelines for the Fresh Tomato Supply Chain*. Second edition. NATTWG and United Fresh Produce Association. https://www.wga.com/sites/wga.com/files/resource/files/12ST TOMATOGUIDE.pdf.

Grace, D. and J. McDermott. 2015. Reducing and Managing Food Scares. In *2014–2015 Global Food Policy Report*, Washington, DC: International Food Policy Research Institute. March 18. 41-50.

Hay, B., R. Stavins, and R. Vieter. eds. 2005. *Environmental Protection and the Social Responsibility of Firms: Perspectives from Law, Economics and Business.* Washington, DC: Resources for the Future. pp. 210.

Hidayat, N.K., P. Glasbergen and A. Offermans. 2015. Sustainability Certification and Palm Oil Smallholders' Livelihood: A Comparison between Scheme Smallholders and Independent Smallholders in Indonesia. *International Food and Agribusiness Management Review* 18(3):25–48.

Jargon, J. and J. Newman. 2016. Chipotle Outbreak Probed. *The Wall Street Journal,* Jan. 7, p. B1.

Kuchler, F. 2015. How Much Does It Matter How Sick You Get? Consumers' Response to Foodborne Disease Outbreaks of Different Severities, USDA, ERS, ERR-193. http://www.ers.usda.gov/publications/err-economic-research-report/err193.aspx.

Leeb, S. 2013. Wal-Mart Fattens Up on Poor America With 25% of U.S. Grocery Sales. *Forbes Magazine*, May 20.

Maertens, M. and J. Swinnen. 2015. Agriculture Trade and Development: A Value Chain Perspective. WTO Working Paper ERSD-2015-4. https://www.wto.org/english/res_ e/reser_ e/ersd201504_e.pdf.

McBride, W., C. Greene, L. Foreman, and M. Ali. 2015. The Profit Potential of Certified Organic Field Crop Production. USDA, ERS, ERR-188, July. http://www.ers. usda.gov/ publications/err-economic-research-report/err188.aspx.

McWilliams, A., and D. Siegel. 2001. Corporate Social Responsibility: A theory of the Firm Perspective. *Academy of Management Review* 26:117-127.

Mueller-Loose, S. and H. Remaud. 2013. Impact of Corporate Social Responsibility Claims on Consumer Food Choice: A Cross Cultural Comparison. *British Food Journal* 115(1): 142-162.

National Research Council. 2011. *Potential consequences of public release of Food Safety Inspection Service establishment specific data.* Committee on a Study of Food Safety and Other Consequences on a Study of Food Safety and Other Consequences of Publishing Establishment-Specific Data. Washington, D.C.: Board on Agriculture and Natural Resources.

Nesheim, M., M. Oria, and P. Tsai Yih, Editors. 2015. *Committee on a Framework for Assessing the Health, Environmental, and Social Effects of the Food System.* Food and Nutrition Board; Board on Agriculture and Natural Resources; Institute of Medicine; National Research Council.

Nidumolu, R., C.K. Prahalad, and M.R. Rangaswami. 2009. Why Sustainability is Now the Key Driver of Innovation. *Harvard Business Review* 87(4):57–64.

Paul, C. J., and D. S. Siegel. 2006. Corporate Social Responsibility and Economic Performance. *Journal of Productivity Analysis* 26:207-211.

Pollan, M. 2006. *The Omnivore's Dilemma: A Natural History of Four Meals.* New York: Penguin Press.

Poppe, K.J., S. Wolfert, C. Verdouw and T. Verwaart. 2013. Information and Communication Technology as a Driver for Change in Agri-Food Chains. *EuroChoices* 12(1):60–65.

Roberts, T. 2015. Food Safety Considerations for Nutrition Science. Academy eBriefings. The New York Academy of Sciences, Feb. 6. www.nyas.org/FoodSafety-eB

Roberts, T. 2013. Lack of information is the root of US foodborne illness risk. *Choices*. www.choicesmagazine.org/magazine/pdf/cmsarticle_300.pdf.

Rock, C. 2014. Evaluation of risk-based water quality sampling strategies for the fresh produce industry. Center for produce safety, Award No. 2014–314. http://www.centerforproducesafety.org/.

Saitone, T., R. Sexton, and D. Sumner. 2015. What Happens When Food Marketers Require Restrictive Farming Practices? *American Journal of Agricultural Economics* 97 (4): 1021-1043. doi:10.1093/ajae/aav021.

Scharff, R. 2015. State estimates for the annual cost of foodborne illness. *Journal of Food Protection* 75(1):123-131. dx.doi.org/10.4315/0362-028X.JFP-14-505.

Stephen, S. P. 2004. The Economics of Corporate Social Responsibility. *Export wise* Winter.

Streeter, D., S. Sonka, and M. Hudson. 1991. Information Technology, Coordination, and Competitiveness in the Food and Agribusiness Sector. *American Journal of Agricultural Economics* 73(5): 1465-1471.

Teillant, A. and R. Laxminarayan. 2015. Economics of Antibiotic Use in U.S. Swine and Poultry Production. *Choices* Quarter 1. http://www.choicesmagazine.org/.

Thatcher, M. K. 2015. Who Owns my Data? American Farm Bureau Federation. http://www.fb.org /issues/bigdata/overview/.

Unilever. 2015. Sustainable Living. https://www.unilever.com/sustainable-living/.

United Nations. 1987. Report of the World Commission on Environment and Development: Our Common Future. Transmitted to the General Assembly as an Annex to document A/42/427. Development and International Co-operation: Environment.

United Nations Environment Programme, 2000: Environment and Trade: A Handbook, published by IISD and available at http://www.iisd.org/trade/handbook/5_4_1.htm.

U.S. Department of Agriculture. 2015. National Organic Grain and Feedstuffs. Agricultural Marketing Service.

van Hoose, N. 2015. Laser Tool Speeds Up Detection of Salmonella in Food. Purdue Alumnus. May/June. www.Purdue.alumni.org.

Wal-Mart. 2015. Sustainability Report. http://corporate.Wal-Mart.com/global-responsibility/sustainability/.

Wiedmann, M. 2014. Validation of geospatial algorithms to predict the prevalence and persistence of pathogens in produce fields to improve GAPs. Center for produce safety.

International Food and Agribusiness Management Review
Special Issue - Volume 19 Issue A, 2016

Big Data and Smallholder Farmers: Big Data Applications in the Agri-Food Supply Chain in Developing Countries

Iuliia Protopop[①a] and Aleksan Shanoyan[b]

[a] *Doctoral Student, University of Nebraska-Lincoln, Department of Agricultural Economics, 316 Filley Hall, Lincoln, NE 68503, USA*

[b] *Assistant Professor, Kansas State University, Department of Agricultural Economics 304 G Waters Hall, Manhattan, KS 66506, USA*

Abstract

The potential of *big data* (BD) applications in agriculture is attracting a growing interest from food and agribusiness industry players, researchers, and policy makers. Possible gains in agricultural productivity and supply chain efficiency from BD-based solutions can help address the challenge of doubling the food supply by 2050. Most of the research in this area evolves around commercial agricultural production in developed countries with relatively limited attention to BD-based solutions focused on smallholder farms in developing countries. This paper provides an overview of the existing and emerging technologies that can potentially enhance the BD application in the agribusiness value chain in developing countries, and presents a discussion of four successful cases of BD applications targeting smallholder producers. This paper also highlights drivers and barriers for smallholder-oriented BD applications in the agri-food supply chain in developing countries and discusses related implications for policy makers, private industry, and NGOs.

Keywords: big data, smallholder farms, agribusiness, ICT, Africa

[①]Corresponding author: Tel: + 1 402.840.1272
Email: I. Protopop: iprotopop@huskers.unl.edu
A. Shanoyan: shanoyan@k-state.edu

Introduction

The potential for *big data* (BD) applications in agriculture is attracting a growing interest from food and agribusiness industry players, researchers, and policy makers. The term BD is broadly defined as "high-volume, high-velocity, and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making."[1] Potential gains in agricultural productivity and supply chain efficiency from BD-based solutions could significantly enhance the ability of global agri-food systems to face the challenge of doubling the food supply by 2050. Most of the research in this area revolves around the BD applications in commercial agricultural production in developed countries. Relatively limited attention is given to the potential of BD-based solutions in the agri-food value chains in developing economies (Kshetri 2014).

The restricted access to resources and markets faced by smallholder farmers in developing countries is often cited as the main barrier for adopting new technologies and developing new capabilities that are necessary for successful BD based solutions (Jack 2013). However, two important factors should be considered when assessing the potential of smallholder farm-oriented BD applications: (1) smallholder farms in developing countries are cultivating significant areas of farmland and are responsible for the majority of the local food supply in the most rapidly growing[2] regions of the world (Salami 2010); (2) the adoption of Information and Communication Technologies (ICT), in particular mobile-cellular and mobile-broadband connection, in developing countries is growing at a remarkably rapid rate. The combination of these two factors provides unprecedented opportunity for mass mobilization and aggregation of information and data on the smallholder farm level in developing economies. This in turn creates an enabling environment for BD applications that can benefit smallholder farmers and enhance the efficiency of agri-food supply chain.

Many new technological solutions are currently transforming a number of key industries in developing countries and creating significant but complex opportunities for BD applications with potential benefits to actors across the agri-food value chain. Therefore, there is a need to explore: 1) how the ICT growth can support the use of BD applications, 2) where and how BD is currently being used in developing countries, and 3) what are the major drivers and impediments of the adoption of BD applications. The review of the existing research in this area revealed a large gap in the academic literature related to BD applications in agriculture in developing countries. The existing documented evidence on this topic is largely limited to technical reports, proof-of-concept studies, and project descriptions. The purpose of this paper is to begin filling the gap by making the existing evidence on smallholder-oriented BD applications easily accessible to food and agribusiness scholars and decision makers.

The objectives of this paper are threefold. First, it provides an overview of the existing and emerging technologies that can potentially enhance the BD application in the agribusiness value chain in developing countries. Second, it presents discussions of successful cases and examples of BD applications in agricultural and related sectors in developing countries. Third, it highlights drivers and barriers for BD application in the agri-food supply chain in developing countries and discusses implications for policy makers, private industry, and NGOs.

---

[1] Gartner IT Glossary 2015. http://www.gartner.com/it-glossary/big-data/.
[2] United Nation Development Programme, World Population Prospects, 2015.

The definition of BD has evolved in the recent literature to better reflect what BD applications allow rather than focusing purely on the dimension of the information asset (volume, velocity, and variety) (Boyd and Crawford 2012, Kshetri 2014, Sonka 2014, Linville 2015). In the context of this study, we combine Gartner and Kshetri's definitions and refer to BD as data sets that (1) are of higher volume[3], (2) are of greater variety[4], (3) provide insights that were not available before, and (4) enable timely decision making for supply chain actors and policy makers. Accordingly, the assessment of BD-based solutions and their potential in developing countries can be based on two alternative approaches. The first approach involves adopting the current use of BD in developed countries as a benchmark, while the second approach entails setting the benchmark at the highest possible potential of BD use in a particular region/country. Different types of benchmarks can lead to very different sets of conclusions in the analysis of current state of, relevance, and barriers to BD use in the developing world.

This paper uses the second approach since it allows conducting a more accurate and realistic assessment without the risk of underestimating the potential of BD applications in developing countries. It presents an overview of BD applications focusing on smallholder producers in developing countries. The analysis is based on the information from a variety of sources including academic literature, policy documents, and popular media. The remainder of the paper is organized as follows: the next section provides an overview of the current state and the potential of the BD-enabling technological infrastructure in developing countries, followed by the presentation of four cases of successful BD applications in East Africa, specifically in Kenya and Uganda, as well as examples from Central America and Southeast Asia. The last section presents a discussion of major drivers and barriers for smallholder farm oriented BD solutions in developing countries and concludes with implications for the agribusiness industry and policy makers.

Big Data-Enabling Infrastructure in Developing Countries

One of the critical conditions for successful BD applications is the presence of capacity and infrastructure for generating, processing, storing, and distributing the large-volume, wide-variety, and high-velocity data. Some common elements of such infrastructure in developed countries include mobile networks, internet and social media, network devices and sensors, and satellite communication systems among others. Many of these elements are either unavailable or are not well developed in developing countries as indicated by the Networked Readiness Index, which measures the performance of 148 economies in leveraging information and communications technologies at a global level to boost competitiveness and well-being. (World Economic Forum 2015) (Figure 1). However, over the last five years an unprecedented growth of ICTs has been observed, in particular mobile-cellular and mobile-broadband subscriptions (Figure 2). While the internet penetration in developing countries remains at a relatively low rate of 35.3% compared to 82.2% in developed countries[5], the rate of increase in the number of new subscribers has been remarkable over the last decade. Between 2007 and 2015, the percent of the population in the developing world using internet increased from 1–40%. Most of the increase in internet use in developing countries is due to the rapid growth in mobile-broadband connection. In 2015, thirty-nine out of 100 people in developing countries had mobile-broadband subscriptions. Between 2005 and 2015, mobile-

[3] Compared to traditional data sets.
[4] Compared to traditional data sets.
[5] ICT Facts and Figures, 2015.

broadband use has increased thirty times. While the large proportion of increase in mobile-broadband use in developing countries is likely driven by city dwellers, the general infrastructure for mobile-broadband service in rural areas is largely present through mobile-cellular coverage (International Telecommunication Union 2014). For example, the percentage of rural population covered by a mobile-cellular signal in developing regions was around 80%, according to 2012 data (Figure 3).

According to 2014 data, 90% of the population in developing countries own and use a cell phone. Data from sub-Saharan Africa indicate that 63% of rural and 80% of urban households own at least one cell phone (Tortora 2013). Thus, even though sensor and satellite technologies are not broadly available in developing regions, the rapid growth of ICTs has a potential to enhance the BD-enabling environment and allow for generation, storage and analysis of large volume and greater variety of real- or near real-time data.

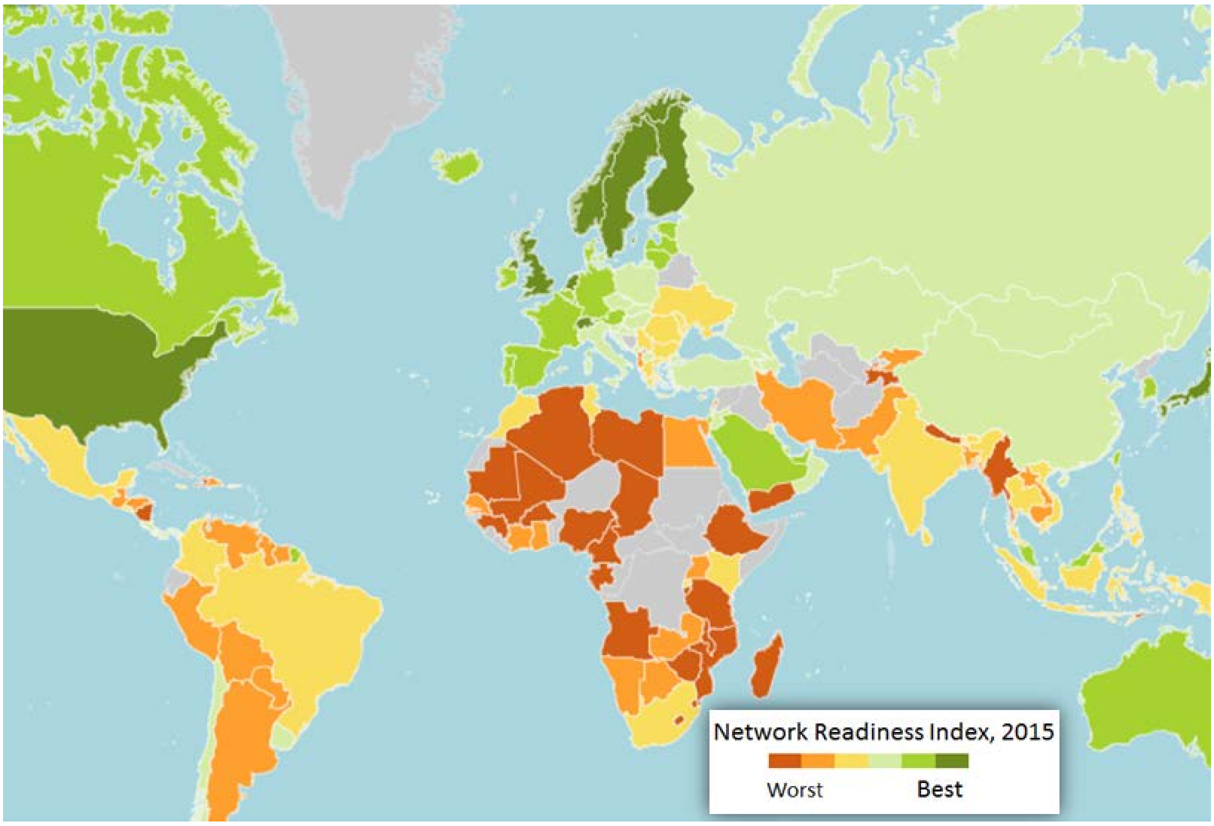


Figure 1. Networked Readiness Index, 2015.

Note. Networked Readiness Index measures, on a scale from 1 (worst) to 7 (best), the performance of 143 economies in leveraging information and communications technologies to boost competitiveness and well-being.
Source. Map by World Economic Forum, Interactive heat map of country/economy profiles, 2015.

Another element of technological infrastructure needed for storage, processing and access of BD is often referred to as *cloud computing and storage*. Cloud computing and storage technology can be simplified into two major components: (1) data centers with large storage capacity and (2) communication capabilities of broadband infrastructure for data transmission and access. The 2013 United Nation's (UN) Information Economy Report on Cloud Computing in Africa identified lack of infrastructure, fixed-broadband connectivity, and power-supply issues as problems that prevent African countries from developing and growing cloud computing. However, external stakeholders see high-growth potential in African markets and the use of shared data storage centers to achieve cost savings from economies of scale. According to the 2013 UN Report on Information Economy, 15% of co-location data centers are based in developing countries. These include seventeen centers in South Africa, one in Nigeria, and two in Kenya.

BD analysis is largely done by third-party private firms who have BD analytics capabilities and offer services to a range of private and public clients. For example, mobile network operators, such as Safaricom, and lending institutions, such as Central Bank of Africa, outsource the data analysis to private analytics firms such as Cignify and Experian for producing consumer-risk profiles. There are also instances of in-house data processing, one such example is Telenor Group - a mobile network operator providing services in thirteen markets including a number of developing countries in Eastern Europe and Asia (Aschim 2014). Most likely the majority of BD analysis will continue to be outsourced to large specialized companies due to the importance of economies of scale in this industry.

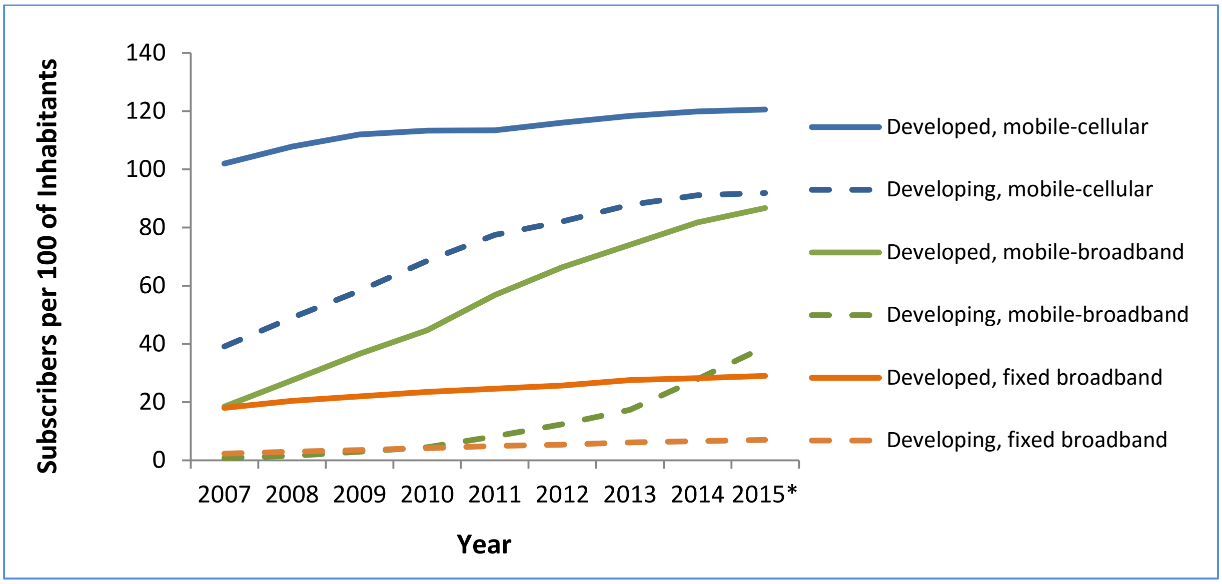


Figure 2. Growth of key ICTs in developed and developing countries, 2007–2015.

Source. International Telecommunication Union (ITU) World Telecommunication/ICT Indicators Database, 2015.
* Projected value, ITU 2015.

The ICT expansion has already had a significant impact on several key industries in developing countries including healthcare, banking, and the public service sector. In fact, in some countries these industries have been practically redefined through ICT and BD applications. For example, a BD analytics company partnered with a Mobile Network Operator (MNO) in Sierra Leone (Airtel SL) to be able to better predict the spread of the Ebola virus. The BD application was based on the analysis of call detail records (CDRs) and signaling data to predict population mobility within and around affected areas to be able to forecast the spread of the virus. This information in turn was used to help health care providers to better prepare and respond to outbreaks and medical emergencies. (Real Impact Analytics 2014). In another example from Tanzania, the BD was used to improve the availability and access to anti-malaria medicine and other critical medication by collecting real-time data via text messaging from health care providers on the current stock levels of medicine. The generated data and the insights from the analysis were used to more accurately forecast the needs and improve the delivery of life-saving medications, reducing re-stocking related incidents from 78% to 26% (Newton 2012).

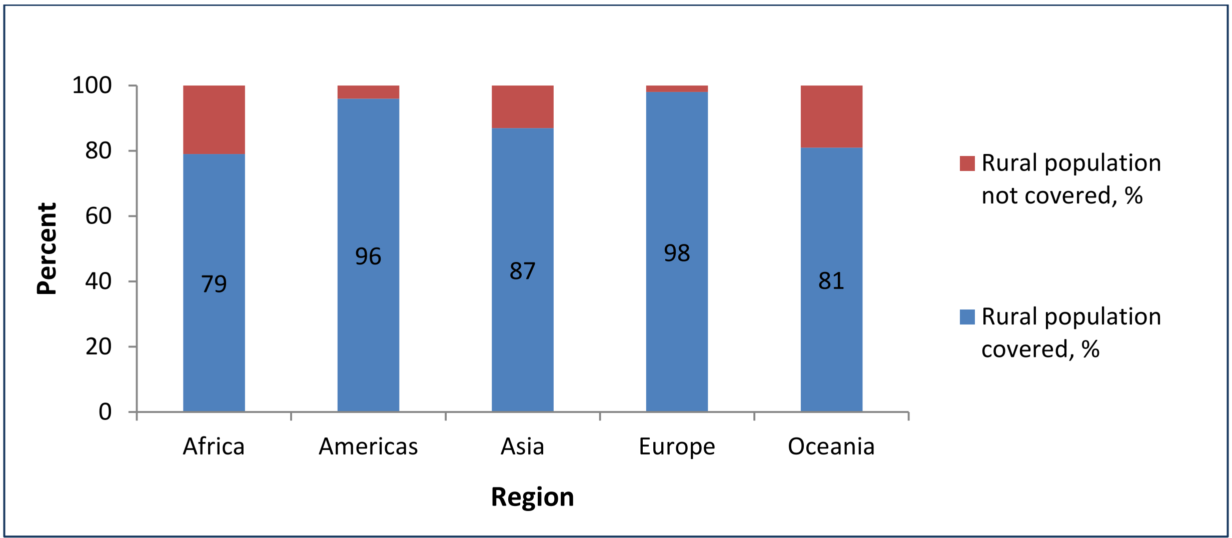


Figure 3. Rural population covered by a mobile-cellular signal, 2012.
Source. International Telecommunication Union Measuring the Information Society Report, 2014.

In Kenya, BD analytics in combination with crowdsourcing technologies allowed generating real-time information on the condition of water wells in rural areas which allowed much faster response to droughts and water shortages in remote areas (Benady n.d.). In the financial service sector, the use of ICT and mobile devices for banking services have directly impacted millions of smallholder producers in Africa by allowing them to make and receive payments with much lower transaction costs, to help better balance their cash flows within the production year, to build a credit record, and to invest in productivity (Demirguc-Kunt 2014). These and other examples of successful ICT enabled BD-based solutions in developing countries can provide important lessons and insights on the potential of BD applications focused on smallholder producers.

Cases of Successful Big Data-Based Solutions for Smallholder Farmers

Index-Based Agriculture Insurance in East Africa: The Case of Kilimo Salama

This case illustrates how BD applications enabled the design and delivery of an innovative agricultural insurance program to smallholder farmers initially in Kenya, and later was extended to Rwanda and Tanzania (International Finance Corporation n.d.). Kenya is a country of 44.9 million people in East Africa with more than 75% of the population making a living by subsistence farming (Salami 2014). Agriculture in this country as in many other developing regions is characterized by large numbers of smallholder farms and continues to be the major source of employment and livelihood (Gollin 2014). Thus, public and private initiatives focused on improving smallholder farm productivity and profitability are widely considered the key for ensuring food security and long-term economic development.

Agricultural policy makers and the agribusiness community have long recognized the importance of effective risk-management tools. This is especially true for Kenya and other developing countries where resource-constrained smallholder farmers are particularly susceptible to adverse weather events and economic shocks. However, many past attempts to design and deliver crop insurance products to smallholder farmers in developing countries, including Kenya, have proven to be economically and operationally inviable. Specifically in

Kenya, the traditional insurance plans designed to cover smallholder farm output required agents to travel to the site to evaluate the condition of the crop and to assess the damage. The combination of poor road infrastructure and the large number of remote farm locations made this system too costly for both the insurer and the farmer (International Finance Corporation, Ideas42 n.d.). Additionally, this type of output-based insurance product created corruption and fraud risks as it was practically impossible to oversee interaction between individual agent and client. Lastly, there were significant trust issues due to the lack of effective enforcement mechanisms and incidents of non-payment of indemnities by insurance companies (Cole 2013). All of these conditions made it difficult for insurance companies to establish lasting relationships with farmers and led to a rapid drop in the participation rates and ultimate failure of traditional insurance programs.

Recent developments in ICT and the widespread adoption of mobile communication in Kenya presented unprecedented opportunity for innovation in the field of agricultural insurance targeted towards smallholder farmers. The opportunity was recognized and seized by UAP Insurance, a large insurance company in Kenya which in partnership with the Safaricom, the largest mobile network operator in the country and Syngenta Foundation for Sustainable Agriculture (SFSA) developed and launched a new agricultural insurance product for smallholder farmers (International Finance Corporation n.d.). The insurance product, called Kilimo Salama[6], was designed to help farmers manage the risks from rainfall variability by covering farmers' inputs rather than outputs and using the data-driven objective index to determine indemnities therefore eliminating the need for traditional subjective evaluation by the loss adjuster (Farming First 2013). Additionally, the partnership with Safaricom allowed UAP and farmers to carry out all transactions/payments via cell phone technology, which significantly reduced the cost of the program delivery, and allowed extending the product to a larger number of smallholder farmers in remote areas.

The insurance company installed satellites and state-of-the-art weather stations in regions where it offered the product to collect and transmit weather data such as rainfall level, wind speed, and temperature to the Kilimo Salama cloud-based server every fifteen minutes (SFSA 2014). The real-time weather data was transmitted to the central location where it was combined with the regional level historical climate and crop yield data and processed using Syngenta's analytical models to estimate indemnities. The minimum rainfall level (also referred to as a "trigger level") was calculated for each specific region using thirty-year climate and crop yields data in that particular region.

The Kilimo Salama pilot program, using two weather stations, was launched in 2009 with 200 farmers who purchased the product. The following year, the participation rate had gone up to 12,000 farmers in a much wider area with twenty-five additional weather stations. In 2013, the product insured 187,467 farmers in three countries (Kenya, Rwanda and Tanzania), and the company plans further expansion in Zimbabwe, Nigeria, Ethiopia, Ghana and Uganda by 2016 (International Finance Corporation n.d.). The product quickly gained farmers' trust because having the weather insurance not only protected them from significant economic losses, but also improved access to credit, encouraged investments in high-productivity inputs, and increased production efficiency. The January 2014 Kilimo Salama's Review shows that farmers who purchased weather insurance invested 20% more in their operations and generated 16% more income than those operators who were not insured. Access to insurance allowed farmers to be qualified for micro-loans (177,782 farmers -or 97% of

[6] Translated from Kiswahili, Kilimo Salama means "Safe Agriculture".

insured operators- received loans totaling $8.4 million) which were not available to them in the past (SFSA 2014).The product was a success for UAP since the company was able to (1) develop a more objective measure due to generating weather data and using it to determine indemnities, (2) reduce the program delivery cost and thus improve supply chain efficiency in delivering this product, and (3) build trust with farmers (International Finance Corporation, Ideas42 n.d., World Farmers' Organization n.d.). Additionally, UAP's insurance served as an indirect coverage for microfinance institutions (MFIs) by protecting their debtors—smallholder farmers. Thus, this BD-based risk protection effectively helped farmers to improve their loan eligibility and their ability to invest in production.

The use of mobile phones to develop and deliver an effective insurance product in Kenya impacted multiple stakeholders. From the insurance company's perspective, this solution enhanced product design which in turn improved efficiency in product delivery and reduced the cost of delivery. From the smallholder farmers' perspective, the new risk management tool helped them to become more financially resilient to weather shocks. In countries like Kenya, the aggregate impact of such a risk management tool can be very high since rural communities are composed of many smallholder farmers (World Bank 2014). This case provides a good illustration of how an ICT enabled risk management solution was made possible through a collaborative effort of various stakeholders and significantly enhanced smallholder agricultural production and food security in the region. While this case does not present a direct application of BD, the ICT-based insurance solution contributes to the BD infrastructure by establishing a capacity for the generation of large volumes of wide variety, real-time data that can be used by multiple stakeholders (such as micro-finance institutions and input suppliers) to enhance efficiency throughout the supply chain.

Big Data Based Solution for Enhancing Financial Inclusion of Smallholder Farmers: the Case of M-Pesa and M-Shwari in Kenya.

This case illustrates BD applications designed to enhance smallholder farmer's access to credit. The application was developed and implemented through a partnership between mobile network operators (MNO), data analytics companies, and financial institutions. The joint initiative made it possible to use individual telecommunication and transaction data to develop a credit scoring system for smallholder farmers in Kenya. According to the World Bank's 2014 Global Findex Database, in 2011 about 2.5 billion people in the world did not have a bank account and thus were unable to save, borrow, or transfer funds outside of cash transactions. The limited access to banking services and credit has been shown to impede smallholder farm's productivity and growth (Foltz 2004).

The recent widespread adoption of mobile communication had a dramatic impact on smallholder banking in sub-Saharan Africa, enhancing transaction efficiency and trade in the agri-food sector (Ogbeide and Ele 2015). Introduction of mobile money accounts expanded financial inclusion providing 64 million people the ability to use banking services for conducting transactions, borrowing funds, and saving. For approximately half of these mobile money account holders, mobile banking is the only way to access banking services because traditional banking services are not available in many remote areas (Demirguc-Kunt 2014).

The new modes of payment through mobile technology not only significantly reduced transaction cost for smallholder farmers but also generated large volumes of data on transactions by millions of individual farmers. The data on millions of smallholder farm sales volume, price, transaction date, frequency, and payment history which was practically impossible to track and record in the past became abundant due to mobile banking. The BD analytics capabilities made it possible to analyze the mobile banking data and to assess individual smallholder farmer's credit risk profiles and borrowing/repayment capacity. This in turn created an unprecedented opportunity for financial institutions to extend banking and credit services to agricultural producers who were credit constrained due to the lack of credit history.

In sub-Saharan African countries, and in Kenya in particular, the mobile network operators were the first to realize the value of extending financial services to the unbanked population and using it as an additional revenue stream. Groupe Speciale Mobile Association (GSMA), the association of mobile operators and related companies, with the financial support from the Bill and Melinda Gates Foundation and Rockefeller Institute, launched the industry-wide initiative called mAgri which focuses on using the available ICTs to enhance the quality of life of smallholder farmers. Two of the program's most important components are Agricultural Value Added Services (AgriVAS) and Agricultural Mobile Finance Services (AgriMFS) (GSMA 2015).

AgriVAS offers a number of agricultural value-added services such as agronomy information and advisory services via mobile phone to farmers in remote areas. The information is transmitted via text messages and the cost of which is billed to customers. The program is well aligned with the goals of the MNOs as it helps them to: (1) retain and enhance loyalty of existing customers by offering new services that help enhance their yields, (2) attract new customers and (3) enhance revenue stream from text messages. As a result, the project began to generate large volumes of data from client call detail records (CDRs) which were matched with demographic information provided at registration. This provided initial BD infrastructure to launch mobile money programs, such as Agri MFS. These programs generated streams of new user-level data on agricultural transactions and a potential for BD application in this area.

Historically, the absence of credit risk profiles and payment history were the major obstacles faced by financial institutions in their efforts to extend credit products to smallholder farmers. However, the new data generated from AgriVAS and AgriMFS provided an unprecedented opportunity to team up with MNOs to design innovative digital financial services focused on smallholder agricultural producers. Among the most notable digital financial services were M-Pesa and M-Shrawi, introduced in Kenya by Safaricom, Kenya's leading mobile network operator. It successfully launched M-Pesa—a mobile money transfer service which is used by more than two-thirds of adult Kenyans, has more than 80,000 agents, and processes $20 million in daily transactions (Cook 2015). The rapid adoption rates created an opportunity for Safaricom to introduce new banking products to the segment of their clients who were previously unbanked but now gained access to major financial services via mobile phone.

M-Shwari was launched in Kenya in November 2012 through a partnership between Central Bank of Africa (CBA) and Safaricom, and offers a combination of savings and loan products. With these new products, clients were able to borrow and repay funds using mobile phones and thus manage their cash flows and savings more efficiently. The M-Shwari is an account managed by CBA, but the transactions are made through the M-Pesa wallet. As of December 2014, M-Shwari extended a total of 20.6 million cumulative loans to their clients and disbursed $277.2 million (Cook and McKay 2015). This product is an example of a BD-based solution used by MNOs and financial institutions in developing countries to accurately determine the loan size and terms of borrowing for smallholder farmers.

The technology used by mobile operators allows for generation of two categories of real time dynamic data: (1) telco, which includes data on demographics, location and mobility, source, and destinations of calls; and (2) payment information, which includes financial information such as payments sent/received, and airtime top-ups. The MNOs collect the data and outsource the analysis to data analytics companies. The results of the analysis are then transferred to the financial institution to make appropriate decisions on loan approval. Three companies—Signify, Experian, and Real Impact Analysis (RIA) - are some of the largest data analytics companies that operate in the developing countries as an intermediary between the lending institutions and the mobile network operators and estimate credit risk profiles for clients using the data generated by MNOs and their predictive models. Recently, some MNOs (e.g., Telenor Group) considered transferring this function in-house to ensure that (1) the data of their clients is secure and protected, and (2) they can access the information at any time. Processing and analyzing BD requires significant investments in resources and capabilities but can provide leverage when negotiating the terms with the financial institutions.

M-Shwari analyzes clients' credit-worthiness and assigns credit limits using an algorithm based on customer use of Safaricom services (Cook 2015). The variables used in the algorithm are telecommunication data including airtime and airtime credit based on the Safaricom's datasets, as well as, the use of M-Pesa and the length of time as an M-Pesa customer. The real-time nature of the data being collected from the clients allows re-evaluating clients' payment performance and adjusting the loan size and terms of borrowing very promptly. The latter in turn makes borrowing more efficient while reducing the probability of a non-performing loan (NPL).

This illustrative case of a BD-based financial solution contributes to the overall objective of this paper in three ways. First, it shows how the growth and development of ICTs enabled the utilization of BD in the banking and financial sector. Second, the case illustrates the ability/potential of BD to develop and deliver products/services to remote smallholder farmers otherwise inaccessible due to poor physical infrastructure and high transaction costs. Third, it shows how the BD-based solution enabled lending institutions to extend credit to smallholder farmers who otherwise would not have access to it due to the absence of individual credit history.

Enhancing Smallholder Farmers Access to Market Information through Big Data Applications: The Case of AgriLife in Kenya and Uganda

The case of AgriLife in East Africa is an example of successful use of a BD-driven platform to improve access to market information and reduce inefficiencies in the agri-food supply chain by connecting multiple stakeholders. Uganda is an east African country with a population of 37.6 million, 80% of which are living in rural areas and deriving their livelihoods from agriculture. The country's economic development has been constrained significantly by many factors including the lack of appropriate transportation and communication infrastructure. Recently, mobile phone technology was rapidly adopted and widely used by Ugandans. In 2014, Uganda's communication commission reported 19,506,550 subscribers of mobile phones which is a 10% increase from 2013. The rapid adoption and widespread use of mobile technology generated large volumes of a wide variety of real-time user level data, opening up a possibility for BD applications in a range of industries including agriculture. The potential value for the agri-food sector from such BD applications was recognized by both public and private sectors leading to the development and launch of several public-private initiatives in this area.

Approximately 47% of Ugandans have access to financial services including mobile-based and traditional banking. The number of Ugandans using mobile money services is four times greater compared to those using exclusively traditional bank accounts. In Kenya, 65% of the population has access to ranking/financial services, more than half of which have only mobile money accounts (InterMedia Financial Insights 2014). The high rate of mobile technology adoption in the communication and banking sector established an initial BD generation and aggregation infrastructure enabling development of a larger integrative platform for connecting millions of smallholder farmers with financial institutions, buyers, input suppliers, and agricultural service providers.

One such platform was developed and launched by Mercy Corps, an international development organization, in Uganda (CTA 2014; Jimenez 2013). With the financial support of Swiss Agency for Development and Cooperation, and in partnership with MobiPay (Kenya-based IT-company), Mercy Corps developed a mobile-based platform called AgriLife. The purpose of AgriLife was to bring all of the stakeholders along the agribusiness supply chain into an integrated data driven system in order to meet smallholder farmers' needs faster and more effectively (MercyCorps 2015). The platform facilitates steps in the BD application process including: (1) collection of near real-time data on farmers' production capability and history, borrowing/repayment potential, and input use, (2) data analysis and projection of future production capacity, demand for inputs and credit, and (3) data integration across supply chain to identify farmers' needs more accurately and deliver resources to distant farmers in a more timely manner (Technical Center for Agriculture and Rural Cooperation 2014). Due to new capabilities offered by this platform, in two years, more than $2 million was extended to about 120,000 distant smallholder farmers in Kenya and Uganda in revolving credit lines (Kshetri 2014).

The developers of AgriLife found building long-lasting trusting relationships with farmers to be one of the key factors in attracting farmers to join the network. They partnered with farmer-centric agribusiness enterprises (e.g. farmer cooperatives, producer associations) to reach out to farmers via entities which already have established trust and reputation with farmers. To be part of the platform, farmers need an active mobile phone and a subscription. Once subscribed, farmers are able to transact with farmer-centric enterprises via mobile phone at significantly lower transaction costs. The farmer-centric enterprises collect data on farmers' production history, input use, etc. via mobile phone and share the data within the platform. MobiPay data analytics experts use the generated data to project production capacity, predict demand for inputs, estimate borrowing capacity and develop credit risk profiles. This information is then used by lenders and input suppliers to supply products/services more efficiently to smallholder farmers.

The use of this platform provides a number of benefits to agri-food supply chain actors. First, by sharing their production, demographic, and transaction data, farmers are able to signal to input suppliers, lenders, and buyers about their production potential and need for credit and other inputs. Second, input suppliers are able to identify and more accurately estimate the demand for their products and services. Third, financial institutions can use the credit scores, payment history, and production data provided via this platform to develop and offer more targeted financial services to the broader customer base. This is an illustration of how the BD application can facilitate transactions in the agri-food supply chain by enhancing access to market information.

Linking Smallholder Farmers in Central America, East Africa, and Southeast Asia to Export Markets through a Cloud-Based Transactional Platform: The Case of FarmForce

The use of various out-grower schemes aiming to connect smallholder fruit and vegetable growers with export markets are becoming a commonly observed phenomenon in developing countries. However, the increasingly stringent global trade protocols with strict traceability requirements have created unfavorable conditions for smallholder farmers who lack necessary resources and capabilities for investing and implementing effective traceability programs (Feed the Future 2014). Consequently, there was a need for an innovative solution for enhancing farmers' ability to meet export market standards and certifications while at the same time ensuring a more stable and predictable supply of good quality produce for exporters (Jimenez 2013). Syngenta Foundation for Sustainable Agriculture -SFSA- (with co-investor, State Secretariat for Economic Affairs of Switzerland, and partners, Global GAP and MercyCorps) provided such a solution through a BD application called FarmForce.

FarmForce is a cloud-based mobile application which allowed farmers and field staff to enter the real-time production data which was then directly transmitted to the exporter's server and analyzed for further management, logistics, and distribution decisions. The platform was first introduced in East Africa (Kenya 2012) and then expanded to Central America (Guatemala, 2013) and Southeast Asia (Farmforce n.d.).

In Guatemala, a private horticulture exporter, Fair Fruit, signed up with FarmForce in 2013 to improve its grower management capabilities. Before having access to this product, the firm experienced difficulty with efficiently collecting information (Global G.A.P. 2014) which could negatively impact smallholders (if their produce does not meet the food safety requirement) and the exporter (if unable to ensure a stable supply of produce). In 2012, Fair Fruit enrolled only 16% of its farmers, but by 2014 it extended subscription to all of its smallholders (FarmForce n.d.). Some of the benefits of using this platform include the improved knowledge of the farmer profile (information on personal profiles and location that was difficult to obtain in the past) and the high efficiency of information transfer between the head office (Fair Fruit) and field staff/farmers (Global G.A.P 2014).

Discussion and Conclusion

In both developed and developing countries BD applications in agriculture are largely driven by the desire to improve productivity and profitability of firms along the agri-food supply chain. However, BD applications in the developed world are primarily focused on enhancing the productivity and efficiency of large-scale commercial agriculture, while BD-based solutions in developing countries have been largely focused on addressing systemic problems and market failures along the agri-food supply chain. Consequently, there are also significant differences in the nature of the opportunities and challenges for BD applications in developed and developing world.

One of the primary drivers of BD-based solutions in developing countries is the rapid advancement of ICTs, specifically decreasing the cost of data storage, growing mobile-cellular coverage in rural areas, and increasing use of mobile phones even among the most resource-constrained smallholder farmers. While the mere collection of smallholder farm data through ICTs would not constitute BD, the data generated through the use of mobile phones is uniquely detailed and in combination with additional supplemental information can produce significant benefits for agri-food stakeholders through ex-post analysis, real-time measurement and feedback, as well as future prediction and planning. For example, the presented cases illustrate how the ex-post analysis of cellphone-generated transaction data can provide otherwise unattainable insights on creditworthiness of millions of individual smallholder farmers. The cases also illustrate how the analysis of real-time spatial and temporal data from transactions using SMS and other value-added services can result in more accurate assessment of access to input and output markets, spread of animal and plant diseases, as well as in the delivery of timely alerts and mitigation assistance. Further, the relatively high inter-comparability of the format of mobile-cellular data produced and held by different telecommunication operators provides a significant scaling potential and the ability to improve accuracy of predictive analysis of crop yields, food supply and demand shocks, and potential food security risks. This is consistent with key features of the BD definition, namely (1) high volume, (2) greater variety, (3) generating previously unavailable insights, and (4) enabling timely decision making by supply chain actors and policy makers.

Another key enabling factor for smallholder-oriented BD applications is the growing number of collaborative win-win solutions mutually beneficial for both public and private sectors. The private firms are the primary drivers of BD applications in the developed world, while in the developing countries many BD solutions are launched through public-private initiatives with the support of NGOs and international development agencies. For example, as highlighted in the Kilimo Salama case, the new agricultural insurance product was developed and extended to farmers through collaboration between the private mobile network operator Safaricom, large insurance company UAP, and a non-profit Syngenta Foundation for Sustainable Agriculture. In the examples of BD based solutions for financial inclusion and market information problems in Kenya and Uganda, the applications were designed and implemented through a collaborative effort of a variety of stakeholders including private MNOs and data analytics firms, micro-credit institutions, and development agencies such as the United Nations and World Bank. With the growing awareness about potential benefits of smallholder-focused BD applications, there is an increasing realization of advantages of public-private partnerships in exploiting that potential in developing countries.

Moving forward, the international development community, governments, and the agri-food industry will have to direct significant effort to overcome a number of important challenges and barriers in order for benefits from smallholder-oriented BD applications to fully materialize. Some of the key barriers include availability and accessibility of data, standardization and interoperability of BD analytics, data privacy and security concerns, as well as underdeveloped legal infrastructure for governing the sector (International Telecommunication Union 2014; Naef et al. 2014). Many specialists in BD strategy who point out the rapid growth of data generation in the developing world also emphasize the importance of data sharing (Rijmenam 2015). According to the World Economic Forum’s report Big Data, Big Impact (2012), emerging markets use a much narrower set of technologies for data collection compared to industrialized countries; however, they are able to generate about 2.5 quintillion bytes of data every day, and the annual growth of mobile-generated data is expected to exceed 100% through 2015. The report also emphasizes the need to establish BD alliances in the developing world to capture gains from synergies in collective capabilities and to tackle the issue of wide digitalization gap between developed and developing countries (World Economic Forum 2012). Sharing data with other stakeholders can create large potential benefits due to network effects and economies of scale and scope. However, it also raises concerns about data privacy and security. Thus, the successful application of BD is conditional on not only the presence of technological infrastructure, but also the adequate legal infrastructure.

This paper provides an overview of the current state, growth potential, and key drivers of the ICT adoption and potential utilization in BD applications in developing countries. The analysis of four illustrative cases of smallholder-oriented BD applications suggest that the rapid advancement of major technologies such as mobile phones and cloud-based technology have a potential to create favorable data collection and aggregation infrastructure for further development of BD applications in developing countries. The stronger focus on ITC in some of the cases is driven by the important enabling role of ICT for BD applications in the context of developing countries. As evidenced by the case studies, this rapidly growing technological infrastructure enables development of BD based solutions to systemic failures in the agri-food supply chain in developing countries. Moving forward, there is a need for more research by agribusiness and development scholars in order to gain insights on drivers and impediments of innovative BD based solutions to study or assess problems faced by smallholder farmers. This is an important gap to fill since the BD can unlock the agricultural production potential of smallholder farmers and significantly enhance food security in regions with low economic development and high population growth.

References

Aschim, Hege. 2014. Leveraging telco data for predictive credit scorecards in mobile money: An interview with Telenor. The GSMA Foundation. Blog. http://www.gsma.com/ mobile fordevelopment/programme/mobile-money/leveraging-telco-data-for-predictive-credit-scorecards-in-mobile-money-an-interview-with-telenor [accessed September 5, 2015].

Boyd, D. and K. Crawford. 2012. Critical questions for Big Data: provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication, and Society* 15(5):662-679.

Cole, S., X. Gine, J. Tobacman, R. Townsend, P. Topalova, and J. Vickery. 2013. Barriers to household risk management: Evidence from India. *American Economic Journal: Applied Economics*. 5(1): 104-135.

Benady, David. n.d. Can big data improve the lives of people in the developing world? *The Guardian*. http://www.theguardian.com/sustainable-business/2014/dec/11/can-big-data-improve-the-lives-of-people-in-the-developing-world [accessed March 5, 2016].

CGIAR. (n.d.). Big data for climate-smart agriculture. CCAFS. Website. https://ccafs.cgiar.org/bigdata#.VuNFjfkrLIU [accessed February 21, 2016].

Cook, T. and C. McKay. 2015. How M-Shwari works: the story so far. *The Consultative Group to Assist the Poor*. http://www.cgap.org/publications/how-m-shwari-works-story-so-far [accessed September 25, 2015].

Cook, Tamara and K. McKay. 2015. Top 10 things to know about M-Shwari. 2-April. *CGAP* http://www.cgap.org/blog/top-10-things-know-about-m-shwari [accessed October 13, 2015].

CTA. 2014. AgriLife: Bringing the agriculture value chain finance ecosystem to life. The Technical Center for Agriculture and Rural Cooperation. Blog.14- July. http://blog.cta.int/2014/07/14/agrilife-bringing-agriculture-value-chain-finance-ecosystem-life/ [accessed October 5, 2015].

Demirguc-Kunt, A., L. Klapper, D. Singer, and P. Van Oudheusden. 2014. The Global Findex Database. Measuring Financial Inclusion around the World. *World Bank, Development Research Group.* https://openknowledge.worldbank.org/bitstream/handle/10986/21865/WPS7255.pdf?sequence=2 [accessed September 20, 2015].

Farmforce. (n.d.). Locations. http://www.farmforce.com/en [accessed June 3, 2016].

Farmforce. (n.d.). Fair-Fruit, Case study. http://www.farmforce.com/assets/adisagua_case_study_en.pdf [accessed March 2, 2016].

Farming First. 2013. Kilimo Salama – Index-based agriculture insurance. 10-September. Website: http://www.farmingfirst.org/2013/09/kilimo-salama-index-based-agriculture-insurance/ [accessed February 20, 2016].

Feed the Future. 2014. Mobile app modernizes agricultural trade in Guatemala. Feed the Future, October 30. http://feedthefuture.gov/article/mobile-app-modernizes-agricultural-trade-guatemala [accessed February 14, 2016].

Foltz, Jeremy. 2004. Credit market access and profitability in Tunisian agriculture. *Agricultural Economics* 30(3): 229-240.

Tortora, Bob. 2013. Africa Continues Going Mobile. *Gallup.* http://www.gallup.com/poll/168797/africa-continues-going-mobile.aspx [November 1, 2015].

Gartner.com. 2015. Big Data. http://www.gartner.com/it-glossary/big-data/ [accessed October 17, 2015].

GLOBAL G.A.P. 2014. Farmforce. New mobile application for smallholders in cooperation with Global G.A.P. 27-June. GLOBAL G.A.P. http://www.globalgap.org/uk_en/media-events/news/articles/New-GLOBALG.A.P.-Mobile-Application-for-Smallholders-in-Cooperation-with-Farmforce/ [accessed March 2, 2016].

Gollin, Douglas. 2014. Smallholder agriculture in Africa: An overview and implications for policy. International Institute for Environment and Development Working Paper. IIED, London.

GSMA. 2015. Mobile for Development. http://www.gsma.com/mobilefordevelopment/ [accessed September 7, 2015].

InterMedia. 2014. Financial Inclusion Insights Quicksights Report: Kenya. http://finclusion.org/wp-content/uploads/2014/04/FII-Kenya-Wave-One-Survey-QuickSights-Report-updated-2014.pdf [accessed September 12, 2015].

InterMedia. 2014. Financial Inclusion Insights Quicksights Report: Uganda. http://finclusion.org/wp-content/uploads/2014/03/FII-Uganda-Wave-1-Public-Quant-QuickSights.pdf [accessed September 12, 2015].

International Finance Corporation. 2015. Kilimo Salama – Index-based Agriculture Insurance: A Product Design Case Study. Washington, DC. International Finance Corporation. https://openknowledge.worldbank.org/handle/10986/21682 [accessed September 18, 2015].

International Finance Corporation. (n.d.). Agriculture and Climate Risk Enterprise – Kilimo Salama – Kenya, Rwanda, Tanzania. World Bank Group. http://www.ifc.org/wps/wcm/connect/industry_ext_content/ifc_external_corporate_site/industries/financial+markets/retail+finance/insurance/agriculture+and+climate+risk+enterprise [accessed February 20, 2016].

International Telecommunication Union. 2014. The World in 2014: ICT facts and figures. https://www.itu.int/net/pressoffice/press_releases/2014/23.aspx [accessed October 2, 2015].

International Telecommunication Union. 2014. Measuring the information society report. https://www.itu.int/en/ITU-D/Statistics/Documents/publications/mis2014/MIS2014_without_Annex_4.pdf [accessed February 17, 2016].

International Telecommunication Union. 2015. The World in 2015: ICT facts and figures. http://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2015.pdf [accessed October 7, 2015].

Jack, B. K. 2013. Constraints on the Adoption of Agricultural Technologies in Developing Countries. Literature review, Agricultural Technology Adoption Initiative, J-PAL (MIT) and CEGA (UC Berkeley).

Jimenez, Andrea. 2013. The Farmforce platform. *E-Agriculture*, May 10. http://www.e-agriculture.org/news/farmforce-platform [accessed March 2, 2016].

Jimenez, Andrea. 2013. AgriLife platform: New mobile service could boost farming in Uganda. *E-Agriculture*, January 30. http://e-agriculture.org/news/uganda-new-mobile-service-could-boost-farming [accessed September 17, 2015].

Kshetri, Nir. 2014. The emerging role of big data in key development issues: opportunities, challenges, and concerns. *Big Data and Society* 2:1-20.

Linville, Charles, Big Data Applied: Market and Transportation Location Analytics, Informa Economics. GeoSilos Big Data Roundtable, May 28, 2015.

MercyCorps. 2015. Unleashing Ugandan farmers' potential through mobile phones. MercyCorps, Global Envision. https://www.mercycorps.org/research-resources/unleashing-ugandan-farmers-potential-through-mobile-phones [accessed September 17, 2015].

Naef, E., P. Muelbert, S. Raza, R. Frederick, J. Kendall and N. Gupta, 2014. Using Mobile Data for Development. Whitepaper. Cartesian and Bill and Melinda Gates Foundation. http://digitalprinciples. org/wp-content/uploads/2015/05/Using-Mobile-Data-for-Development.pdf . [accessed February 25, 2016].

Newton, Alice. 2012. Big data for development: Beyond Transparency. The World Bank Blog. 23-July. http://blogs.worldbank.org/psd/big-data-for-development-beyond-transparency [accessed March 5, 2016].

Ogbeide, Osadebamwen A. and I. Ele. 2015. Smallholder farmers and mobile phone technology in Sub-Sahara Agriculture. *Mayfair Journal of Information and Technology Management in Agriculture.* 1(1): 1-19. https://www.academia.edu/9786327/Smallholder_Farmers_and_Mobile_Phone_Technology_in_Sub-Sahara_Agriculture [accessed February 3, 2016].

Rijmenam, Mark. 2015. How big data can help the developing world beat poverty. https://datafloq.com/read/big-data-developing-world-beat-poverty/168 [accessed September 10, 2015].

Salami, A., A. Kamara, and Z. Brixiova. 2010. Smallholder Agriculture in East Africa: Trends, Constraints, and Opportunities. Working Paper Series #105. Africain Development Bank. http:/www.afdb.org/ [accessed July 5, 2015].

Sonka, Steve. 2014. Big Data and the Ag Sector: More than Lots of Numbers. *International Food and Agribusiness Management Review* 17(1): 1–20.

Syngenta Foundation for Sustainable Agriculture. 2014. “Kilimo Salama” is ACRE. http://www.syngentafoundation.org/index.cfm?pageID=562 [accessed February 15, 2016].

Syngenta Foundation for Sustainable Agriculture. (n.d.). Kilimo Salama Annual Review. http://www.syngentafoundation.org/__temp/Kilimo_Salama_3_Pager_21_1_14.pdf [accessed February 21, 2016].

Uganda Communications Commission. 2014. Facts and Figures. http://www.ucc.co.ug/data/qmenu/3/Facts-and-Figures.html [accessed October 14, 2015].

United Nations Conference on Trade and Development. 2013. *Information Economy Report: The Cloud Economy and Developing Countries.* Geneva: United Nations Conference on Trade and Development. http://unctad.org/en/PublicationsLibrary/ier2013_en.pdf [accessed September 15, 2015].

United Nations Global Pulse. 2012. *Big Data for Development: Challenges and Opportunities*. New York: UN Global Pulse. http://www.unglobalpulse.org/ sites/default/files/BigDataforDevelopment-GlobalPulseMay2012.pdf [accessed at August 25, 2015].

United Nations Development Programme. 2015. *World Population Prospects: The 2015 Revision. Key Findings and Advance Tables.* New York: UN Population Division. http://esa.un.org/unpd/wpp/Publications/Files/Key_Findings_WPP_2015.pdf [accessed at September 27, 2015].

World Bank. 2014. Rural Population, database. http://data.worldbank.org/indicator/SP.RUR.TOTL [accessed September 12, 2015].

World Economic Forum. 2015. *The Global Information Technology Report 2015: ICTs for Inclusive Growth.* Geneva: World Economic Forum. http://reports.weforum.org/global-information-technology-report-2015/report-highlights/ [accessed on October 30, 2015].

World Economic Forum. 2012. Big Data, Big Impact: New Possibilities for International Development. Geneva: World Economic Forum. http://www3.weforum.org/docs/WEF_TC_MFS_BigDataBigImpact_Briefing_2012.pdf [accessed October 12, 2015].

World Farmers' Organization. (n.d.). Kilimo Salama (Safe Agriculture) – Index-based agriculture insurance. http://www.wfo-oma.com/climate-change/case-studies/kilimo-salama-safe-agriculture-index-based-agriculture-insurance.html [accessed March 9, 2016].

International Food and Agribusiness Management Review
Special Issue - Volume 19 Issue A, 2016

Enhancing Food Safety, Product Quality, and Value-Added in Food Supply Chains Using Whole-Chain Traceability

Brian D. Adam[ⓘ][a], Rodney Holcomb[b], Michael Buser[c], Blayne Mayfield[d], Johnson Thomas[e], Corliss A. O'Bryan[f], Philip Crandall[g], Dar Knipe[h] Richard Knipe[i] and Steven C. Ricke[j]

[a] *Professor,* [b]*Professor, Department of Agricultural Economics, Oklahoma State University, 413 Ag Hall and 114 Food & Agricultural Products Center, Stillwater, OK 74078, USA*

[c] *Associate Professor, Department of Biosystems and Agricultural Engineering, Oklahoma State University, 113 Agricultural Hall, Stillwater, OK 74078, USA*

[d] *Associate Professor,* [e]*Associate Professor, Department of Computer Science, Oklahoma State University, 232 MSCS, Stillwater, OK 74078, USA*

[f]*Post-Doctoral Associate,* [g] *Professor,* [j] *Professor, Department of Food Science and Center for Food Safety, University of Arkansas, 2650 N. Young Avenue, Fayetteville, AR 72704, USA*

[h, i] *Food Systems SME, MarketMaker / Riverside Research, 4709 44th St., Suite 7, Rock Island, IL 61201, USA*

Abstract

A robust whole chain traceability system can limit consumers' exposure to potentially hazardous foods, improve supply chain management, and add value to consumer products. However, fragmented supply chains present special challenges. In the beef industry, for example, producers have resisted participation in whole chain traceability because of high cost relative to value and concerns about disclosing proprietary information, among others. A multi-disciplinary team from universities, private firms, and a foundation has developed and tested a pilot proprietary centralized data whole chain traceability system that addresses many of these obstacles. This system would facilitate a precision agriculture approach to beef production and marketing. While the remaining challenges are serious, the benefits to society, consumers, and businesses from widespread adoption of whole-chain traceability systems are potentially large.

Keywords: whole chain traceability, fragmented supply chains, MarketMaker®, proprietary centralized data, food safety, product recall, value-added, product tracking, biosecurity, big data

[ⓘ]Corresponding author: Tel: + 1.405.744.6854
Email: B. D. Adam: Brian.Adam@okstate.edu

Introduction and Motivation

A robust whole-chain traceability system can provide the foundation for a targeted and timely product recall after a food-borne illness outbreak. It can limit consumers' exposure to potentially hazardous foods and strictly limit a company's liability.

In the beef industry, one of the best-known food-borne illness outbreak examples is the 1992–1993 incidences of Escherichia coli O157:H7 illnesses stemming from undercooked hamburgers served in Jack-in-the-Box restaurants. Foodmaker, Inc., owner of Jack in the Box, issued a recall but ultimately only recovered 20% of the potentially contaminated meat. In the aftermath, Foodmaker lost approximately $160 million in sales and 30% of its stock market value, and subsequently paid tens of millions of dollars to settle individual and class-action lawsuits (Soeder 1993). The CDC conducted a traceback in an attempt to discover the source of the contaminated beef. The CDC identified six separate slaughter plants (five in the US and one in Canada) as the likely sources of the contaminated ground beef (CDC 1993). Animals slaughtered in US plants were further traced to farms and auction houses in six states. The CDC was not able to identify a specific slaughter plant nor farm associated with the contaminated beef. If there had been a robust whole-chain traceability system in place it might have limited human suffering and financial losses.

Another example illustrating the difficulty of beef recalls in the absence of full traceability capabilities is the XL Foods beef recall. In 2012 it was discovered that as much as 2.5 million pounds of beef product involved in the recall issued by the Canadian company had entered the United States and had been distributed in at least eight states (Goetz 2012). The recall was further complicated by the possibility that companies could have used the recalled product to produce other products such as ground beef, ground beef patties, beef jerky or pastrami. Eventually recalls were issued for steaks, roasts, and ground beef products from US retailers including but not limited to Safeway, Sam's Club, Walmart, Albertson's, Fred Meyer, and Kroger (Bottemiller 2012). This incident resulted in the sale of the plant identified as the source of the contamination and payment of millions of dollars to settle lawsuits (Food Safety News 2016). An independent review concluded in part that XL Foods was not prepared to handle a large-recall multi-country incident (Lewis et al. 2013).

A whole chain traceability system would allow sources of contamination in the supply chain to be identified and unsafe food recalled because information could be traced end to end (McKean 2001; Smyth and Phillips 2002). Although the Bioterrorism Act of 2002 requires one-up, one-down traceability, a firm-by-firm traceback in the event of a food safety or bioterrorism event is inherently slow, even with good records at each supply chain node. As part of the Food Safety Modernization Act of 2011 (FSMA), each step in the food supply chain is required to keep its records in digital form in addition to or instead of paper to make the records more accessible to government officials. However, this change in itself is likely to only modestly increase the speed of what would still be a firm-by-firm traceback. A robust whole-chain traceability system is needed that facilitates rapid information transfer up and down the supply chain.

The technology for such a system is available. Golan et al. (2003) noted more than a decade ago that "…retailers such as Walmart and Target have created proprietary supply-chain information systems that their suppliers must adopt" (p.17), observing, moreover, that these are not just for packaged products but for the flow of raw agricultural inputs and outputs. The authors suggested that the private sector has significant capacity for tracing, with incentives to trace for food safety and quality control, for differentiating and marketing foods with credence attributes, and supply chain management. Systems such as these are typically in vertically integrated or tightly controlled supply chains. Many vertically-integrated supply chains are able to effectively trace backward and track forward because information flows within the same company.

When a supply chain is fragmented, though, transactions occur across several companies or continents, and technological and institutional constraints make tracing and tracking products exceedingly more difficult. In order to achieve whole-chain traceability, a firm at one stage needs to share information with the firm at the next stage and in turn through each firm/stage of the supply chain. The more information that is shared, the better the supply chain coordination—but the greater the risk that the information can be used by competitors. As Crandall et al. (2013) noted, deep-seated concerns by firms of disclosing their proprietary information is a key obstacle to implementing whole-chain traceability systems in fragmented supply chains. Other obstacles include perceptions that cost of implementation and operation are greater than value, lack of standards for sharing information, and potential for increased liability. These obstacles have severely limited potential participation by firms in whole-chain traceability systems, thus greatly limiting value of traceability for improving food safety and improving value to consumers. If few firms participate, even technologically-advanced traceability systems accomplish little – data that is not collected cannot be used.

Because many of these obstacles to adoption of whole-chain traceability have been observed clearly in the beef supply chain, leading to abandonment of the National Animal Identification System (Schroeder and Tonsor 2012), this article:

1) highlights key benefits and challenges of implementing a whole-chain traceability system for beef supply chains;
2) describes a technology designed (a proprietary centralized data whole-chain traceability system, or PCD-WCTS) to address the identified obstacles and challenges of implementing a whole-chain traceability system in fragmented supply chains; and
3) identifies remaining challenges for implementing whole-chain traceability systems in fragmented supply chains.

Although the specific application described here is to beef cattle supply chains, the PCD-WCTS technology is readily adaptable to other supply chains. To illustrate the potential for adding value to other food supply chains, the article describes an interface between PCD-WCTS and a private company's traceability system (which allows two-way communication between producers and consumers), as well as a planned value-added interface between PCD-WCTS and MarketMaker®, a web-based platform for matching buyers and sellers of a range of food products and commodities.

Food Safety, Big Data, and Whole Chain Traceability

Big Data analytics is the process of examining large data sets containing a variety of data types to uncover hidden patterns, correlations, trends, customer preferences and other useful information. The use of such data, including data generated by a WCTS, offers great potential for improving food safety. The CDC estimated that there are 47.8 million cases of foodborne illness in the United States and more than 3,000 affected persons die every year (Scallan et al. 2011). The World Health Organization (WHO) estimated that worldwide there are 2.2 million deaths every year from diarrheal food- and water-borne illness, with almost 90% of those deaths being children (WHO 2015). These are staggering statistics of the burden of food- and water-borne illness, but they are largely preventable with currently available technology.

Large data sets now exist that provide the opportunity to change from a mindset where the majority of our food safety resources are focused on routine testing to an informed mindset where the limited resources of regulators and the food industry can be targeted in a proactive manner to minimize the risks of occurrence of food borne illness. Noteworthy examples include:

> whole genome sequencing, which permits more rapid identification of disease-causing pathogens and matching with specific sources (Dumitrescu, Dauwaldera, and Linaa 2011; Köser et al. 2012; den Bakker et al. 2015; Inns et al. 2015; Orsi et al. 2008; and Stasiewicz et al. 2015); real-time internet searches of social media, which may allow speedier identification of illness outbreaks (Grein et al. 2000; Heymann and Rodier 2001; Wilson et al. 2008; and Wilson and Brownstein 2009); and Geographic Information Systems (GIS), which can help identify potential problems and adjust production to reduce those problems, and help predict outbreaks using location and weather data (Scallan et al. 2011; Beuchat 2006; Fremaux et al. 2008; and Johnson et al. 2003).

Accordingly, beginning with HACCP (Hazard Analysis, Critical Control Points) in the late 1990s, traceability and information management have become central to both voluntary and mandated food safety efforts. The principles of both the food industry's Global Food Safety Initiative (GFSI) and FSMA are undeniably tied to whole-chain traceability and data management to support traceability. As Golan et al (2003, 18) stated, "(p)roduct-tracing systems are essential for food safety and quality control."

Using big data analytics, a well-functioning WCTS could capture producer data and interface with other data generated along the supply chain, including consumer data. This large data repository could be used for analytics and visualization. Data analytics algorithms could be used to analyze anonymized data from producers, processors, distributors, vendors, and consumers to create an early animal disease, food safety, and bioterrorism detection system. Along with disparate data sources, they could be used for prediction (predicting food outbreaks), clustering (identifying clusters of outbreaks), associations and correlations (associating food outbreaks with environmental or other conditions), classification (classifying the extent and seriousness of the outbreak), optimization (based on customer preferences or actions during an outbreak), sentiment analysis (determining the sentiments of customers at different stages of the outbreak), to name a few possible applications for food traceability. This can result in faster, better decision-making, facilitating more effective and quicker responses to food outbreaks.

In addition, data visualization can be used to place the data in a visual context to help people understand the significance of the data. Visualization-based data discovery tools allow users to mash up disparate data sources to create custom analytical views. These tools will support creation of charts and maps as well as interactive, animated graphics on desktops and mobile devices. Such tools can visualize the extent of an outbreak in the form of a map, the rate of spread of the outbreak in the form of an animated graphic, and the populations affected as a chart.

WCTS data may also be exported to decision-based systems such as a beef productivity system that could recommend feed rations to improve meat tenderness, and data on carcass and growth performance of progeny of individual cows could flow back upstream, allowing for improved management of those cows. The data can be used to refine estimates of expected progeny differences (EPDs) for important traits, ultimately improving value added to consumers.

While the importance of whole-chain traceability systems is widely recognized, the food industry has also recognized the challenges of implementing them. Fritz and Schiefer (2009) aptly stated "the establishment of tracking and tracing capabilities meets many barriers that have prevented their broad based use beyond what is legally required."

Consumer responses to food safety concerns, their food safety expectations, and willingness to pay for food safety attributes/practices have been acknowledged in numerous studies (e.g. Bitsch, Kokovic, and Rombach 2014; Lim et al 2013; Yeung and Lee 2012). However, it is also important to understand how consumers perceive the shared responsibility for maintaining food safety standards. A 2010 survey by the International Food Information Council (IFIC 2010) found that consumers place the responsibility for ensuring food safety on all active and regulatory participants in the food marketing chain, but primarily government (identified by 74% of respondents), food manufacturers (70%), farmers/ producers (56%), and retailers/food service (49%).

Ng and Salin (2012) noted that "food safety is an inherently complex agribusiness problem (p.22)," and that safety of a final product is determined by all the production, handling, processing, and retailing practices by all firms involved. Their institutional model helps explain how management decisions by each firm can achieve competitive goals such as profitability and market share, while still achieving the public's need for food safety. In such a complex system, traceability is only one tool in promoting a safer food system. However, it is a vital tool, as noted by IFT's (2009) technical report, which found that all fifty-eight companies in their study sample "…acknowledged the importance of an effective (rapid and precise) product tracing system in safe guarding their supply chain (p.2)."

Whole chain traceability systems can be extremely complex, especially in the case of processed foods. In processed foods, different lots of various raw materials are combined into several production batches that are distributed to numerous points of sale (Hu et al. 2013). Thus, processors must record data not only on the product but also on the processes that impact the product, such as transport, storage and sales (Kim et al. 1995). Traceability systems must support both tracking and tracing, where tracking follows a product along the supply chain with records being kept at each stage, while tracing is the reverse process (Thakur and Hurburgh 2009).

Golan et al. (2003) argued that even without mandated traceability, firms in the United States have several motives for establishing traceability systems and, as a result, private-sector traceability systems are extensive. They suggested that firms establish product tracing systems in order to: 1) improve supply-side management; 2) differentiate and market foods with subtle or undetectable quality attributes; and 3) facilitate traceback for food safety and quality. Widespread adoption of an interconnected WCTS could provide even more valuable information specifically suited for tracking food safety events, tracing them quickly back to their source(s), and even predicting events further down the supply chain (Bhatt et al. 2013; Golan et al. 2003; Golan et al. 2004).

There are numerous traceability systems in the US within vertically integrated companies. However, the majority of the vertically integrated companies with traceability systems share limited or no information with outside companies unless ordered by a court. This is generally to protect information that companies view as critical in maintaining their market share. The challenge is even greater in fragmented supply chains, in which products pass from one stage to another, often with a change in ownership. Even large buyers face input supply chains with stages that are difficult to link together in a traceable manner. The comprehensive traceability review of the seafood industry by Sterling et al. (2015) recognized this, noting that internal traceability systems that allow companies to trace within their own operations were common in the seafood industry, but that the ability to trace transactions from firm to firm was much less common.Sterling et al. (2015) highlighted important traceability success and profitability determinants, noting the irony that the more important and imbedded traceability is in a businesses' operations, the more challenging it is for them to quantify its value. This may help explain why some firms do not see enough value in traceability to adopt it. In fact, in their recommendations section the authors noted that a significant portion of the seafood industry is made up of fragmented supply chains, and that the majority of these businesses did not see value in traceability. Sterling et al. (2015) called for future research to help those businesses "…better understand how traceability helps manage risk, reduce costs, and increase relative competitive position" (p.241).

Benefits of a Whole-Chain Traceability System in the Beef Supply Chain

Implementation of a WCTS potentially brings several benefits to a company, an entire industry, and society. Both domestic and foreign purchasers of a company's or US food products can have increased confidence in the safety of those products, increasing demand for them. This can improve sales and profitability for the industry (Sterling et al. 2015; USDA-AFIS 2009). However, an especially valuable benefit of implementing a WCTS in a supply chain may be increasing value added to consumers and to other supply chain participants. Traceability systems are commonly used within a vertically integrated or tightly coordinated supply chain for quality control and other benefits of supply chain management (Golan et al. 2003; Golan et al. 2004), including, in the case of livestock, improving animal disease traceability.

Whole chain traceability can also improve supply chain management in fragmented supply chains (Sterling et al. 2015). In the case of livestock, Schroeder and Tonsor (2012) observed that animal identification (ID) and traceability systems have developed rapidly around the world, and that most major beef export countries have created animal traceability systems to better protect animal health and to enhance export market growth. Tonsor and Schroeder (2006) noted that an example of a successful whole-chain traceability system for beef is Australia's Traceability and Meat Standards Program and National Livestock Identification System (NLIS). In that system, 99.5 percent of movement transactions are electronically recorded within twenty-four hours of the transaction. It is claimed that the NLIS has created market opportunities for Australia's beef industry amounting to hundreds of millions of dollars (AUD), partly because the NLIS has increased the perception by importers that Australian beef is dependable (VCM International).

The Australian system is also estimated to have led to $200 million (AUD) in net benefits since its introduction by improving value added to consumers. Specifically, by tracking and comparing cattle performance, consumers were statistically more likely to have a "more pleasurable eating experience" (VCM International 2014, 15). Although the Australian NLIS has reportedly been very successful, it has come in part through government mandates following a food safety event that resulted in a quarantine of Australian beef and in part because Australia's heavy dependence on exports led to greater motivation by industry participants (Tonsor and Schroeder 2006).

Estimates suggest that the US beef industry also would experience positive results if it adopted whole-chain traceability. Some of the most comprehensive economic assessments of the value of a national animal identification and traceability have focused most heavily on the role of such a system in avoiding the large costs of reduced exports in the event of an animal disease event. Schroeder and Tonsor (2012) summarized the considerable efforts of the livestock industry and US government agencies in an attempt to establish a National Animal Identification System (NAIS), efforts that gained traction with the 2004 discoveries of BSE ("mad cow disease") in Canada and the United States, but that ultimately were abandoned in 2010.

One might ask why the US did not mandate traceability as did other countries such as Australia. The answer to that question is beyond the scope and purpose of this article, but others have dealt with this issue. Goldsmith et al. (2003), for example, identified historical and political differences leading to differences between European Union and United States approaches to food safety regulation, and developed an institutional model to help understand variations across food safety policy environments.

Ortega and Peel (2010) noted that since animal health programs are part a broader set of human health and food safety systems, there is a public nature to animal ID programs (a basic form of traceability), and that this strengthens the argument for a mandatory system, as well as public investment in such a system. They also observed, though, that implementation of an animal ID system has been politically difficult in many countries for social and cultural reasons, but also because of multidimensional factors affecting costs and benefits, which make it difficult for producers to fully value the uncertain benefits of animal ID relative to its certain costs.

Here, we assume that political realities are such that a traceability system for beef will not be implemented by mandates alone, so the focus of the technology is to reduce costs and increase benefits to individuals to make participation attractive. One could also view increasing the economic attractiveness of traceability as reducing the combined economic and political barriers to participating, which might mitigate negative response to mandated traceability.

To show the potential value of a traceability system for the beef industry, Schroeder and Tonsor (2012) cited estimates by Coffey et al. (2005) that the beef industry had lost $3.2 billion to $4.7 billion in just one year, 2004, due to export restrictions alone after the BSE discoveries. Similarly, Pendell et al. (2010) estimated that if lack of traceability resulted in at least 25% of beef product being unacceptable in international trade, the US could lose a total consumer and producer surplus of $6.65 billion. Viewing those results from another perspective, a 1% increase in domestic demand or 34.1% increase in export demand would fully cover the cost and surplus loss of adopting a traceability system that achieved a 90% participation rate. Even with a 70% participation rate, the research showed that there would be a net benefit to producers of $9.26/head (NAIS 2009). Thus, overall industry benefits would be quite high relative to costs of implementation. These results supported earlier results by Resende-Filho and Buhr (2006) that showed the positive impacts of traceability in the beef and pork industry when substantial negative food safety news is reported by the media, by comparing revenue under the assumption the country has adopted a beef and pork National Animal Identification System with an assumption of no traceability system.

The USDA-APHIS (2009) report on NAIS highlighted other key benefits of an effective animal identification and traceability system, including the ability to establish containment areas to restore market access, increased transparency and reduced information asymmetry in the supply chain, improved value added efficiency, and enhancement of animal welfare in response to natural disasters. That report did not provide dollar estimates of these benefits. Indeed, as Sterling et al. (2015) noted, benefits to individual firms of participating in a traceability system are inherently difficult to calculate.

The USDA-APHIS (2009) report also noted that countries importing beef are increasingly adopting animal traceability systems for their domestic production, and that such systems are becoming requirements for market access. The report suggests that the United States lags behind world standards for animal ID and traceability, and that without traceability the US would face future challenges in maintaining or increasing beef exports.

From the perspective of value to consumers, studies by Lee et al. (2011), Loureiro and Umberger (2004), Angulo and Gil (2007), and Dickinson and Bailey (2005) showed that consumers on average are willing to pay some premium for traceable beef products. These benefits can be partially transferred to producers. However, few studies have been conducted on the benefits of traceability to the producers who would have to pay for traceability.

Challenges in Implementing a Whole-Chain Traceability System in the Beef Supply Chain

Golan et al. (2004) note that even though society or an industry as a whole would benefit more than the costs of implementing a traceability system that does not necessarily imply that individual supply chain participants would receive a net benefit. This is especially true in fragmented supply chains (Sterling et al. 2015; Bhatt et al. 2013). Seyoum et al. (2013) expanded on beef industry research by Blasi et al. (2009) and Butler et al. (2008), estimating that most of the costs of implementing a WCTS would fall on cow/calf producers, the first link in the supply chain but also the smallest producers. Conversely, most of the benefits would accrue to larger producers and processors further down the supply chain. This result confirmed the perceptions of some producers that costs would be greater for those that could least afford them.

Specifically, Seyoum et al. (2013) estimated that the costs to an individual producer, including the costs of RFID eartags, installing the eartags, and amortized costs of RFID readers (but not including costs of the overall system) range from $5.95/head for small cow/calf producers to $0.41/head for cattle feeders with more than 8,000 head. Costs for small cow/calf producers are fourteen times larger than for large cattle feeders because, as the first stage in the supply chain, they pay for the RFID eartag and its installation, and because fixed costs of RFID readers and other equipment are spread over fewer animals.

Benefits, on the other hand, are more likely to be realized by processors and downstream producers. For example, any premium for tender beef would be received by processors, and gains from improved feeding efficiency would be realized by cattle feeders, even if the higher value originates from improved genetics provided by cow/calf producers. Producers contributing the increased value who are one or more stages removed from the stage at which the benefits are realized are less likely to be rewarded for those contributions. Thus, those producers who bear the biggest proportion of the cost of traceability are also likely to receive the least benefit. This illustrates part of the problem with fragmented supply chains, and explains part of the reluctance of many beef producers to participate in the NAIS. It also explains the conclusion by Schroeder and Tonsor (2012) that existing voluntary traceability systems for beef offer producers the option to target export market opportunities, but that to capture those opportunities the entire vertical supply chain from cow-calf producer through exporter must be closely vertically aligned.

As previously noted, although a National Animal Identification System likely would have generated societal benefits far above its costs (Schroeder and Tonsor 2012), when USDA attempted to implement the NAIS in the mid-2000s, many producers resisted, partly because of this perceived cost inequity but also because they did not want to reveal proprietary information that could be used against them by competitors or government agencies, and they believed the costs exceeded the benefits. (Schroeder and Tonsor 2012; Crandall et al. 2013; Adam et al. 2015). These and other factors led to abandonment of the NAIS efforts in 2010 (Schroeder and Tonsor 2012).

Producers participating in a WCTS also face an increased liability risk. In the absence of a WCTS, a food safety event might be traced back to a processor, who bears the cost of recalls and lawsuits. The ability to trace the source of an event back to individual producers, while potentially improving food safety in the supply chain, exposes those producers to risk that they would not face if they did not participate in a WCTS (Golan et al. 2004; Pouliot and Sumner 2008; Crandall et al. 2013).

Implementation costs are composed of cultural, sociological, political, and economic components. Some of these may be actual, quantifiable costs, while others may be based on perception. Ultimately, the key to implementing voluntary WCTS in fragmented supply chains is that incentives must exceed costs for all supply-chain participants.

PCD-WCTS Technology – One Proposed Solution to Resolve Traceability Issues in Fragmented Supply Chains

USDA’s National Integrated Food Safety Initiative (NIFSI) funded a multi-institution, multidisciplinary research project to address these fragmented supply chain issues by developing a pilot scale proprietary centralized data whole chain traceability system (PCD-WCTS) technology for beef cattle. The fundamental design criteria included: 1) stakeholder feedback incorporated into the design; 2) proprietary data, in which entities that enter data into the system, own the data, and control access to that data; 3) centralized data, for greater system integrity and data security; 4) data immutability, in which all records are immutable once an animal or product changes ownership; 5) system is internet based; 6) integrated traceability and product marketing; and 7) system adaptability to non-beef products. The following sections describe the beef cattle pilot-scale PCD-WCTS technology and how it could potentially help resolve obstacles in implementing WCTS in fragmented supply chains, and then highlight remaining challenges for effectively implementing a national-scale version of the technology.

The key advantage of PCD-WCTS, compared to previously attempted and current systems, is one of the level of data access control—the parties putting information into the system maintain granular privacy control over their data. In other words, users putting data into the system decide both who can see that information and what pieces (granules) of information they can see. This is critical, since the ability to trace food through a supply chain depends on private firms sharing product information with competitors as well as collaborators. Moreover, this would address the necessity (noted by Schroeder and Tonsor 2012) of tightly-controlled supply chains for capturing value-added opportunities.

PCD-WCTS as a Data Management System - Features and Capabilities

At the heart of PCD-WCTS is a DBMS (DataBase Management System) that provides a secure filing system for data contributed by PCD-WCTS stakeholders. PCD-WCTS data is housed in a MySQL database located on a Linux server. As a data management system, it is designed to interface with a range of other data input mechanisms. For example, an app for the Apple iPad family of devices has been developed that permits users to access their accounts.

A core component of the system is the database mapping module. This module facilitates interfacing PCD-WCTS with other traceability systems. For example, the pilot version of PCD-WCTS directly interfaces with the traceability system operated by Top 10 Produce LLC d/b/a Beefy Boys Jerky Co. Top 10's system permits producers (its current clients produce oranges, avocados, strawberries, and other fruits and vegetables, as well as beef cattle) to share photos of themselves and their farm, information about how the products were grown, recipes, and any other information the producer believes consumers might want. Food-conscious consumers can view this information simply by scanning the QR code on the product at the grocery store with their smartphones. Consumers can in turn provide real-time feedback to the producers about the quality of the product, or ask questions about the product and how it was raised.

The interface with PCD-WCTS permits Top 10 to extend its relatively short supply chain both downstream and upstream through multiple stages, expanding the number of participants that can access its system even if they are several stages removed. Similar interfaces could be developed with any other traceability systems that need the data management features of PCD-WCTS, as long as the product can be identified digitally.

Ability to Selectively Share Specific Data

Since data stored in PCD-WCTS is owned by the contributing stakeholders, it is important to provide those stakeholders the ability to specify the extent to which their data is visible to others. Once data are in the PCD-WCTS, a second iPad app function permits data-owner users who have stored data in PCD-WCTS to select what data they wish to share and with whom. In this way, a supply-chain participant can assign viewing rights for specific data pieces to specific individuals. This precludes un-authorized viewing, and allows data owners (those who put the data in) complete control over their information. Only firms who have entered information into the system, or those they have pre-authorized—such as other producers, feedlots, and processors—can access that information. The PCD-WCTS has standard and user-defined data-sharing templates. The templates provide the user the ability to define the specific data fields within animal records that would be shared giving users the ability to more rapidly share specific data for specific animals to other specific users. For a detailed explanation and illustration of this feature, see Adam et al. 2015.

Data Immutability

It is critical that the validity of data stored in any system be trusted by its stakeholders. Thus, one of the primary concepts built into PCD-WCTS is that of data immutability. This simply means that existing data values become fixed and unchangeable after certain events take place, such as transferring an animal from one owner to another. As an analogy, consider a contract between two parties. Once the contract has been entered, it generally is not modified; instead, changes or corrections are entered as addenda—or attachments—to the original document. For example, assume that after an animal has been transferred from a producer to a feed lot, it is discovered that the original birthdate of the animal entered into the system is incorrect. Rather than changing the birthdate (which now is immutable because of the transfer), a correction record will be attached to the original record correcting the incorrect information. Thus, both the original, incorrect data as well as the corrected data are available, giving a more complete, trustworthy history of the animal.

Security and the PCD-WCTS Architecture

One of the key decisions to be made in implementing a whole-chain traceability system is the kind of database architecture used. The choice reflects a tradeoff between robustness of the traceability system and perceived independence of each firm (which could affect participation). There are currently at least two possible kinds of architectures for the WCTS database for food; each has committed advocates. One approach is a distributed peer-to-peer model where the database is distributed across multiple sites (Özsu et al. 2011). Each site is, for the most part, self-sustained, managing its own security such as Domain Authentication Services and Application security as well as applications. Each site also manages its own backups, controls its own Internet access, and hosts its own shared files. This is similar to the architecture described by and anticipated for use by the Global Food Traceability Center (GFTC 2014, 6).

A second approach is a centralized (Kroenke et al. 2014) or silo, model. In this approach, all the data is stored in centralized servers. Security, backups, Internet access, shared files, applications are all managed locally at one location. In this architecture, if a specific data request is made and data-owner grants the request, the data would be released from the centralized servers.

Each approach has advantages and disadvantages. An advantage of the distributed approach is that it is more scalable. It also may be perceived as allowing each participating firm more independence, perhaps encouraging greater participation. However, there are disadvantages. In the distributed model there are no uniform system-wide security or backup policies. Each site decides its own security and backup policies. The distributed architecture requires inter-site communications to trace a particular product, so the weakest link in the chain determines the security of the whole system. Under the stress of a product recall, an outage at one firm could break the traceability chain. If a traceback is needed because of a food safety event, gaining access to the needed information depends on each site having its data accessible; the traceback will be only as fast as the slowest link.

In a centralized model, there is separation of data for management and security reasons. In contrast to a distributed model, security, backup and other controls are managed centrally. A disadvantage of a centralized system is that if an attacker breaks or penetrates the security of a centralized silo, he may be able to compromise the whole system. Similarly, a natural disaster could cripple the data server, but this could be mitigated using distributed backup centers. However, there are several important advantages. It is easier in a centralized system to ensure that the database server is stored in a secure server room. Uniform security and backup policies that take into consideration all the stakeholders' requirements can be more easily implemented since all data is stored in secured data centers. There is also less administrative overhead since there is one set of policies. There is full control over potential risk areas such as internet access and there is no need for inter-site communications.

The PCD-WCTS is designed using centralized architecture. Thus, the system incorporates Carestream's (2011) four components of data security: availability, confidentiality, integrity, and tracking ability. Although the system is currently at a pilot scale, preliminary development planning conducted as part of the USDA-NIFSI project developing this system suggests that a scaled-up system could be fully funded with a charge to supply chain participants of 1/2¢ per transaction. The following discussion draws examples from beef supply chains to highlight benefits of using the PCD-WCTS in a fragmented supply chain.

Benefits of PCD-WCTS in a Fragmented Supply Chain

Animal Disease Traceability

USDA-APHIS notes on its website that:

> "Animal disease traceability, or knowing where diseased and at-risk animals are, where they've been, and when, is very important to ensure a rapid response when animal disease events take place. An efficient and accurate animal disease traceability system helps reduce the number of animals involved in an investigation, reduces the time needed to respond, and decreases the cost to producers and the government."

Widespread adoption of a PCD-WCTS in the beef supply chain could greatly expedite a USDA-APHIS investigation, since data observations in the PCD-WCTS are associated with specific animals or products. Because the majority of the data observations are expected to be uploaded and stored in the centralized server, these observations could be analyzed rapidly (assuming the data-owner has given the agency access to the data) to provide timely food safety and animal disease results and projections. Since those who put information into PCD-WCTS own and control it, USDA and FDA are viewed as potential users of the data, much as other participants in the supply chain. A prior arrangement could be made in which producers choose to grant them assess through a template that specifies release of very basic information such as the animal id, time, and location to the USDA on condition that an animal disease event is declared.

Value Added

More and more companies have realized the benefit of using a traceability system to improve supply chain management or to transfer credence attributes along the supply chain. Because PCD-WCTS permits users to control their own data, they have the ability as well as the incentive to use it for a much greater range of value-added purposes.

Value Added to Consumers

In addition to the many supply-side benefits of WCTS use, traceability can be viewed as an assurance of quality and/or safety – a value-added factor in the eyes of consumers. The rise in demand for short supply chain (e.g. local foods) offerings and the successes of MarketMaker® and the Top 10 Produce/Our Locale "know your farmer, know your food" traceability system are anecdotal examples of the value consumers attribute to a food product's traceability for origin or protocol verification.

Deselnicu et al. (2013) confirmed the value of traceability to origin in a meta-analysis of geography-based food valuation studies. The authors concluded that premiums for origin-based labels tend to be greatest in low/no-processed foods with distinct geographic indications, even after accounting for differences across food characteristics and trademarks/brands. Lim et al. (2013), using various models, determined that consumers were willing to pay a premium of $5.85/lb. for traceable beef steaks.

Traceability as a value-added measure of quality assurance and food safety can be directly tied to marketing efforts. Yeung and Lee (2012) demonstrated how marketers can use traceability, quality assurance, and independent organization endorsements as marketing strategies to improve consumers' purchase intentions when food safety concerns exist. The authors found that food industry members can benefit from using trace-based information to assuage consumer anxiety in times of food safety uncertainty.

As an example of value-added to consumers, genetic information is one of many attributes that can be transferred along a chain. DeVuyst et al. (2007) and Weaber and Lusk (2010) noted that certain genetic characteristics have a higher likelihood of resulting in more tender beef cuts. Lusk et al. (2001) found that consumers were willing to pay a premium averaging $1.23/lb. for a tender steak versus a tough one ($1.84/lb. if they were given more information about the steak's tenderness), with 20% willing to pay $2.67/lb. or more.

However, in a typical fragmented supply chain without traceability, it is difficult to convert consumer willingness to pay for desired characteristics into compensation to producers of those characteristics, because supply chains are complex, with many transactions involving products with multiple quality characteristics. Thus, in the beef supply chain, even though consumers are willing to pay more for it, producers receive very little price incentive to provide animals that produce more tender meat. If producers could receive a price incentive large enough to cover additional production costs, they could profitably increase value added to consumers. The PCD-WCTS permits processors to direct premiums as incentives to those producers who provide the increased value, without diluting those incentives by dispersing them through the entire supply chain. The potential value added compares favorably to the estimated ½ cent-per-transaction cost of running the traceability system, noted above.

In this vein, Ge (2014) showed that both producers and processors would benefit economically if they used a whole-chain traceability system in the beef supply chain to provide more tender beef, as one example. If a WCTS were in place that could transfer incentives from processors to cow/calf producers directly to produce animals with genetics favoring more tender meat, results indicate that producers could increase profit per head by $45, considerably more than the approximately $6/head cost of implementing traceability. The net benefits would be higher for producers taking advantage of more than one value-added opportunity (such as improved location of injection sites or providing production information to livestock feed companies). These benefits would depend on an effective system in which information and financial remuneration can be transferred directly from the beneficiaries at one stage to those providing the value, often several stages up the supply chain, rather than through each stage sequentially. Individual producers would not necessarily benefit from adding value to their products unless such a mechanism were instituted.

Value Added to Other Supply Chain Participants (Supply Chain Management)

WCTS can be an especially important tool for applying "precision agriculture" to animal agriculture. The technology allows for data on carcass and growth performance of progeny of individual cows to flow back upstream, allowing for improved management of those cows. Analyzing the collection of *big data* will improve confidence in (EPDs) for important traits. For example, by including information about sires in the data flow of commercial cattle, breed associations could more quickly isolate genetics with superior feed efficiency or tenderness, while assisting cattle feeders in determining optimum time on feed.

Cattle Feeding Efficiency

Feeding cattle is one of the major activities of cattle production. The cost incurred for feeding cattle is the single largest variable cost (Sherman, Nkrumah, and Moore 2010). A traceability system can provide information to improve cattle feed efficiency, providing cost savings. Many feed efficiency genetic characteristics of cattle are moderately heritable (Herring 2003; Elzo et al. 2009). Thus, the cow/calf operator and seed stocker operator could produce cattle with higher feed efficiency through selection or other genetic related management activities. By using PCD-WCTS, this information could be transmitted through a traceability system, from those who provide the value directly to those who can benefit from it and in turn compensate the providers.

In addition to genetic information used to select particular animals and not others, a traceability system can help a feedlot operator optimize the feeding operation by allowing the operator to sort animals by particular characteristics related to feed efficiency, including genetic information. Or, vaccination records for each animal can prevent overmedication of individual animals, reducing costs and the potential for development of antibiotic resistance. In effect, as more information is provided, the more each animal can be treated as an individual and optimal care can be provided. This is especially the case for information that is not readily observable—such as vaccination history and weaning age and weight—as the cattle enter the feedlot but that could be transmitted through a traceability system much more quickly and less costly than with a paper-based system. The PCD-WCTS permits producers to provide this information directly to those who find it valuable, without sharing it with others in the supply chain.

Value Added - An Interface with MarketMaker®

One of the features of PCD-WCTS is its ability to interface with other systems. Even when producers have products with value-added characteristics, participating in a traceability system is not sufficient to give producers access to markets which value those characteristics. Farmers must be able to identify and make their products available to buyers who want value-added attributes. Building an interface between PCD-WCTS and MarketMaker® provides an opportunity especially for small and mid-sized farmers to expand into more differentiated, higher value markets. It would also allow farms and businesses to compete in markets demanding traceability and source verification.

MarketMaker® is a web-based platform that assembles, standardizes and geocodes information on farms and food related businesses in the US. It was initially developed by the University of Illinois as resource for the development of alternative food supply chains organized around marketable points of differentiation. Now supported by land grant universities and state departments of agriculture in more than twenty states, farms and business across the country provide profiles that can be mapped and queried by customers based on specific characteristics. This allows food buyers to identify potential regional and local sources of products with specific characteristics, and allows for more agile coordination of alternative supply chains.

MarketMaker® is currently developing a National Beef Portal to expand farmer/rancher profiles and search parameters to include all farm supplies, production, transportation, and marketing for all beef industry related categories. This would provide a delivery system for more sophisticated business and marketing tools, enhancing value-added capabilities.

Interfacing with PCD-WCTS would provide Portal users the ability to track individual animals through the supply chain, making animals with value-added characteristics visible to MarketMaker® users. For example, cattle feeders, local processors, and even beef marketing firms such as Certified Angus Beef (CAB) or US Premium Beef using MarketMaker® would benefit from using PCD-WCTS to track cattle that have not been implanted or been exposed to antibiotics, or that have the genetic potential to be high marbling. MarketMaker® facilitates matching supply chains for products with value-added characteristics with those desiring those characteristics, so it would help processors, restaurants, and other buyers locate cattle that meet their specifications, including source verification and management and production practices that are identified through PCD-WCTS.

The interface could expand beyond beef, into fruits and vegetables. The interface between PCD-WCTS and *Top 10*, which extends traceability in producer supply chains to the consumer, could be used to aggregate product among producers within the MarketMaker® website. This would provide small producers with the necessary scale as well as with the traceability they need to sell into larger wholesale and retail markets.

Remaining Issues and Challenges

Technology such as the PCD-WCTS can help resolve many of the issues that have hindered widespread adoption of a WCTS in the beef industry. The potential food safety and animal disease mitigation, as well as value added, benefits are large. However, several challenges remain.

Protection of Proprietary Information

Since one of beef producers' main concerns leading to abandonment of the NAIS was lack of maintaining confidentiality of proprietary information, a key feature of the PCD-WCTS is that those who input information control the release of that information. However, unless proper diligence is exercised in setting up the legal framework, it is conceivable that data put into the PCD-WCTS could be subject to Freedom of Information Act requests, or the state-level equivalents. This would discourage participation. While a decentralized architecture might make such litigation more difficult than in a centralized architecture (because of the greater number of potential defendants with a decentralized architecture), it is likely that the legal principles, and the need to address them, would be similar. Legal arrangements that can potentially resolve this issue must be investigated before an appropriate institution is selected to host the data servers and administer the system.

Although the system's key feature of allowing those who input information to selectively share that information should lead to greater producer participation, it is possible that the number of producers choosing to participate and share basic information will not be sufficient to adequately trace animal disease or food safety events. Further research is needed to determine factors necessary to achieve critical adoption and use rates, including determining optimal fees, incentives, and subsidies.

Risk and Liability Re-Allocation from Processors to Producers

Producers participating in a WCTS face an increased liability risk. If an animal disease or food safety event can be traced further up the supply chain to a producer or group of producers, rather than just to a processor, those producers will face increased liability risk. Producers likely have less ability to manage that risk than most processors. While the trace-back ability may increase overall food safety, the reallocation of risk toward producers is likely to dissuade them from participating in a WCTS. In order to lessen the costs of risk reallocation, an indemnification, or insurance, system may be needed.

Transition from Paper to Digital Records

Among the remaining challenges to the implementation of whole chain traceability for the purposes of food safety is the need to convince small producers and manufacturers to transition from handwritten data to digital records. This will require investment in information systems and solutions, including data analysis and training. As part of the process there must be better means to predict a problem before it happens, to become truly proactive rather than reactive. Partnerships could be created that facilitate the use of big data in food safety as well as food production, processing, and distribution. The most important motivator, though, is likely to be as producers and others in the supply chain begin to see that the benefits of both converting to digital records and participating in a WCTS. An interface with MarketMaker® could provide additional training opportunities for producers as well as increase their value-added capabilities.

Large Data Sets in Food Safety Analysis

Another challenge likely will be the cost of analyzing the big data sets related to foodborne illnesses. Armbruster and MacDonell (2014) doubt that agriculture and food industries will be amenable to sharing the cost of developing the specialized skills needed to take advantage of big data. This may lead to more consolidation in the supply chain, as when Monsanto acquired Climate Corporation so that they could have access to localized weather forecasts based on historical data which had been generated while developing insurance proposals to farmers for weather related catastrophes (Bennett 2014). Monsanto is therefore able to offer farmers methods to increase yields by precise timing of field treatments such as fertilization or pesticide applications (Armbruster and MacDonell 2014).

Final Comment

Whole-chain traceability systems can use the information gathered at each stage or node along a supply chain to improve food safety and supply chain management, limit the negative impacts of food safety and animal disease events, and create value-added opportunities for supply chain entities. Fragmented supply chains pose special challenges. Firms sharing proprietary information throughout the supply chain risk having others exploit that information. Moreover, they may not be rewarded for providing information that is valued by a firm several stages up or down the supply chain. A key feature of the technology described here—the ability by firms to selectively share specific data—resolves much of this disincentive for firms to participate in whole-chain traceability. While the remaining institutional challenges are significant, the benefits to society, consumers, and businesses from widespread adoption of whole-chain traceability systems are potentially large.

References

Adam, B. D., M. Buser, B. Mayfield, J. Thomas, C. A. O'Bryan, and P. Crandall. 2015. Computer systems for whole-chain traceability in beef production systems. In *Food Safety: Emerging Issues, Technologies and Systems*. Edited by S.C. Ricke, J.R. Donaldson, and C.A. Phillips. Chapter 2: 9–16 Oxford, UK: Academic Press, Elsevier, Inc.

Angulo, A.M., and J.M. Gil. 2007. Risk perception and consumer willingness to pay for certified beef in Spain. *Food Quality and Preference* 18:1106–1117.

APHIS-USDA. 2009. Overview Report of the Benefit-Cost Analysis of the National Animal Identification System. April (slightly revised). https://www.aphis.usda.gov/traceability/downloads/Benefit_Cost_Analysis.pdf. [accessed February 2016].

Armbruster, W.J., M.M. MacDonell. 2014. Informatics to support international food safety. Proceedings of the 28th EnviroInfo 2014 Conference, Oldenburg, Germany September 10-12, 2014. http://enviroinfo.eu/sites/default/files/pdfs/vol8514/0127.pdf [accessed October 12, 2015].

Bennett, D. 2014. Inside Monsanto, America's third-most-hated company. *Bloomberg BusinessWeek-Technology.* http://www.bloomberg.com/news/articles/2014-07-03/gmo-factory-monsantos-high-tech-plans-to-feed-the-world. [accessed October 12, 2015].

Beuchat, L.R. 2006. Vectors and conditions for preharvest contamination of fruits and vegetables with pathogens capable of causing enteric diseases. *British Food Journal* 108(1):38–53.

Bhatt, T., G Buckley, J. McEntire, P. Lothian, B. Sterling, and C. Hickey. 2013. Making Traceability Work across the Entire Food Chain. *Journal of Food Science* 78(2):B21–B27.

Bitsch, V., N. Kokovic, and M. Rombach. 2014. Risk Communication and Market Effects during Foodborne Illnesses: A Comparative Case Study of Bacterial Outbreaks in the U.S. and in Germany. *International Food and Agribusiness Management Review* 17(3):97–114.

Blasi, D., G. Brester, C. Crosby, K. Dhuyvetter, J. Freeborn, D. Pendell, T. Schroeder, G. Smith, J. Stroade, and G. Tonsor. 2009. Benefit-Cost Analysis of the National Animal Identification System. Final Report submitted to USDA-APHIS on January 14. https://www.aphis.usda.gov/traceability/downloads/Benefit_Cost_Analysis.pdf.

Bottemiller, H. 2012. More illnesses tied to XL Foods beef. http://www.foodsafetynews.com/ 2012/10/more-illnesses-tied-to-xl-foods-beef-recall-expands-to-hong-kong/#. Vs3hN eYTx8F [accessed February 24, 2016].

Butler, L.J., F. Haque, J.W. Oltjen, G. Caja, J. Evans, V. Velez, L. Bennett, and C. Li. 2008. Benefits and Costs of Implementing an Animal Identification and Traceability System in California –Beef, Dairy, Sheep and Goats. University of California Research Report.

Carestream Health. 2011. How to evaluate the data security capabilities of cloud-based services. White Paper. http://www.slideshare.net/Carestream/how-to-evaluate-the-data-security-capabilities-of-cloudbased-services.

CDC. 1993. Update: multistate outbreak of *Escherichia coli* O157:H7 infections from hamburgers - Western United States, 1992–1993. Morbidity and Mortality Weekly Report. 42: 258–263.

CDC. 2008. Outbreak of Salmonella serotype Saintpaul infections associated with multiple raw produce items—United States, 2008. MMWR 57: 929-934.

Coffey, B., J. Mintert, S. Fox, T. Schroeder, L. Valentin. 2005. The Economic Impact of BSE on the US Beef Industry: Product Value Losses, Regulatory Costs, and Consumer Reactions. Kansas State University Agricultural Experimental Station and Cooperative Extension. Serial No. MF-2678, April.

Crandall, P. G., C. O'Bryan, D. Babu, N. Jarvis, M. L. Davis, M. Buser, B. Adam, J. Marcy, and S. C. Ricke. 2013. Whole-chain traceability, is it possible to trace your hamburger to a particular steer, a U. S. perspective? *Meat Science*. 95(2):137–144.

den Bakker, H.C., M. W. Allard, D. Bopp, E.W. Brown, J. Fontana, Z. Iqbal, A. Kinney, R . Limberger, K. A. Musser, M. Shudt, E. Strain, M. Wiedmann, and W. J. Wolfgang. 2014. Rapid Whole-Genome Sequencing for Surveillance of *Salmonella enterica* Serovar Enteritidis. Emerging Infectious Diseases 20(8):1306–1314. doi:10.3201/ eid2008.131399.

Deselnicu, O.C., M. Costanigro, D.M. Souza-Monteiro, and D. Thilmany McFadden. 2013. A Meta-Analysis of Geographical Indication Food Valuation Studies: What Drives the Premium for Origin-Based Labels? *Journal of Agricultural and Resource Economics* 38(2):204–219.

DeVuyst, E. A., J. R. Bullinger, M. L. Bauer, P. T. Berg, and D. M. Larson. 2007. An economic analysis of genetic information: Leptin genotyping in fed cattle. *Journal of Agricultural and Resource Economics* 32(2): 291–305.

Dickinson, D.L., and D. Bailey. 2005. Experimental Evidence on Willingness to Pay for Red Meat Traceability in the United States, Canada, the United Kingdom, and Japan. *Journal of Agricultural and Applied Economics* 37(3):537–548.

Dumitrescu, O., O. Dauwaldera, G. Linaa. 2011. Present and future automation in bacteriology. *Clinical Microbiology and Infection* 17(5): 649–650. doi: 10.1111/ j.1469-0691.2011.03511.

Elzo, M.A., D.G. Riley, G.R. Hansen, D.D. Johnson, R.O. Myer, S.W. Coleman, C.C. Chase, J.G. Wasdin, and J.D. Driver. 2009. Effect of Breed Composition on Phenotypic Residual Feed Intake and Growth in Angus, Brahman, and Crossbred Cattle. *Journal of Animal Science* 87(12):3877–3886.

FDA. 2011. Food Safety Modernization Act (FSMA). http://www.fda.gov/Food/foodSafety/ FSMA/ucm247546.htm [accessed March 1, 2013].

Food Safety News. 2016. Court OKs $4 million class action settlement with XL Foods. http://www.foodsafetynews.com/2016/02/court-oks-4-million-class-action-settlement-with-xl-foods/#.Vs3uL-YTx8F. [accessed February, 2016].

Fremaux, B., C. Prigent-Combaret, C. Vernozy-Rozand. 2008. Long-term survival of Shiga toxin-producing *Escherichia coli* in cattle effluents and environment: an updated review. *Veterinary Microbiology* 132(1–2):1–18. doi: 10.1016/j.vetmic.2008.05.015.

Fritz, M., and G. Schiefer. 2009. Tracking, Tracing, and Business Process Interests in Food Commodities: A Multi-level Decision Complexity. *International Journal of Production Economics* 117(2):317–329.

Ge, Candi. 2014. The Value of a Whole-Chain Traceability System in Transmitting Genetic Information about Beef Tenderness. M.S. Thesis, Oklahoma State University.

Global Food Traceability Center (GFTC). 2014. Concept for an Interoperable Architecture for Seafood Traceability. Discussion Paper. October 30, 2014.

Goetz, G. 2012. USDA confirms recalled Canadian beef entered U.S. *Food Safety News* 21-September. http://www.foodsafetynews.com/2012/09/usda-confirms-recalled-canadian-beef-entered-us/#.Vs3cneYTx8E. [accessed February 24, 2016].

Golan, E., B. Krissoff, F. Kuchler, K. Nelson, G. Price, and L. Calvin. 2003. Traceability in the US Food Supply: Dead End or Superhighway? *Choices.* Second Quarter. (2 2003a):17–20.

Golan, E., B. Krissoff, F. Kuchler, L. Calvin, K. Nelson, and G. Price. 2004. Traceability in the U.S. Food Supply: Economic Theory and Industry Studies. U.S. Department of Agriculture-Economic Research Service, Agricultural Economic Report No. 830. http://www.ers.usda.gov/publications/aer-agricultural-economic-report/aer830.aspx [accessed October 2, 2015.]

Goldsmith, P., N. Turan, and H. Gow. 2003. Food Safety in the Meat Industry: A Regulatory Quagmire. *International Food and Agribusiness Management Review* 6(1):25–37.

Grein, T.W., K.B. Kamara, G. Rodier, A.J. Plant, P. Bovier, M.J. Ryan, T. Ohyama, and D. L. Heymann. 2000. Rumors of disease in the global village: outbreak verification. *Emerging Infectious Diseases* 6(2): 97–102.

Herring, W.O. 2003. Genetics of Feed Efficiency in Angus Cattle. University of Florida.

Heymann, D.L., and G. R. Rodier. 2001. Hot spots in a wired world: WHO surveillance of emerging and re-emerging infectious diseases. *Lancet Infectious Diseases* 1(5):345–353.

Hu, J., X. Zhang, L.M. Moga, M. Neculita. 2013. Modeling and implementation of the vegetable supply chain traceability system. *Food Control* 30(1): 341–353.

IFT. 2009. Institute of Food Technologists/Food and Drug Administration, Contract No. 223-04-2503, Task Order No. 6, Traceability (Product Tracing) in Food Systems, Technical Report, revised. Volume 1 of 2.

Inns, T., C. Lane, T. Peters, T. Dallman, C. Chatt, N. McFarland, P. Crook, T. Bishop, J. Edge, J. Hawker, R. Elson, K. Neal, G.K. Adak, P. Cleary. Outbreak Control Team. 2015. A multi-country *Salmonella* Enteritidis phage type 14b outbreak associated with eggs from a German producer: 'near real-time' application of whole genome sequencing and food chain investigations, United Kingdom, May to September 2014. *Eurosurveillance* 20(16): 15–22.

International Food Information Council (IFIC). 2010. 2010 Food & Health Survey: Consumer Attitudes Toward Food Safety, Nutrition, & Health. IFIC Foundation publication. http://www.foodinsight.org/2010_Food_Health_ Survey_ Consumer_ Attitudes_Toward_Food_Safety_Nutrition_Health#sthash.m2S53ubj.dpbs [accessed October 9, 2015.]

Johnson, J.Y.M., J. E. Thomas, T. A. Graham, I. Townshend, J. Byrne, L.B. Selinger, V.P. J. Gannon. 2003. Prevalence of Escherichia coli O157:H7 and Salmonella spp. in surface waters of southern Alberta and its relation to manure sources. *Canadian Journal of Microbiology* 49(5): 326–335.

Kim, H.M., M.S. Fox, M. Gruninger. 1995. An ontology of quality for enterprise modelling. http://www.eil.utoronto.ca/wp-content/uploads/enterprise-modelling/papers/Kim-WETICE95.pdf [accessed October 9, 2015].

Köser, C.U., M. J. Ellington, E. J. P. Cartwright, S. H. Gillespie, N.M. Brown, M. Farrington, M. T. G. Holden, G. Dougan, S. D. Bentley, J. Parkhill, S. J. Peacock 2012. Routine use of microbial whole genome sequencing in diagnostic and public health microbiology. *PLoS Pathogens* 8(8): e1002824. doi:10.1371/journal.ppat.1002824.

Kroenke, D. and D. Auer. 2014. Database processing – Fundamentals, Design and Implementation. 13th Edition. Upper Saddle River, NJ: Pearson Education, Inc.

Lee, J.Y., D.B. Han, R.M. Nayga, and S.S. Lim. 2011. Valuing traceability of imported beef in Korea: an experimental auction approach. *Australian Journal of Agricultural and Resource Economics* 55(3):360–373.

Lewis, R.J., A. Corriveau, and W.R. Usborne. 2013. Independent review of XL Foods Inc. beef recall 2012. http://www.foodsafety.gc.ca/english/xl_reprt-rapprte.asp. [accessed February 24, 2016].

Lim, K.H., W. Hu, L.J. Maynard, and E. Goddard. 2013. U.S. Consumers' Preference and Willingness to Pay for Country-of-Origin-Labeled Beef Steak and Food Safety Enhancements. *Canadian Journal of Agricultural Economics* 61(1):93–118.

Loureiro, M.L. and W.J. Umberger. 2007. A Choice Experiment Model for Beef: What US Consumer Responses tell us about Relative Preferences for Food Safety, Country-of-Origin Labeling and Traceability. *Food Policy* 32(4):496–514.

Lusk, J. L., J.A. Fox, T. C. Schroeder, J. Mintert, and M. Koohmaraie. 2001. In-store valuation of steak tenderness. *American Journal of Agricultural Economics* 83 (3) : 539–550.

McKean, J. D. 2001. The importance of traceability for public health and consumer protection. *Reviews in Science and Technology* 20(2):363–371.

Ng, Desmond, and Victoria Salin. 2012. An Institutional Approach to the Examination of Food Safety. *International Food and Agribusiness Management Review* 15(2):21–46.

Orsi, R.H., M.L. Borowsky, P. Lauer, S.K. Young, C. Nusbaum, J.E. Galagan, B.W. Birren, R.A. Ivy, Q. Sun, L.M. Graves, B. Swaminathan, M. Wiedmann. 2008. Short-term genome evolution of *Listeria monocytogenes* in a non-controlled environment. *BMC Genomics* 13(9):539. doi: 10.1186/1471-2164-9-539.

Ortega, Carlos, and Derrell S. Peel. The Mexican Animal Identification System: Current Situation, Problems, and Potential. *Journal of Agricultural and Applied Economics* 42(3):551–557.

Pouliot, S., and D.A. Sumner. 2008. Traceability, liability, and incentives for food safety and quality. *American Journal of Agricultural Economics*. 90(1):15–27. doi: 10.1111/ j.1467-8276.2007.01061.x.

Özsu M. A., and Valduriez, P. 2011. *Principles of Distributed Databases.* Third Edition. New York: Springer. doi:10.1007/978-1-4419-8834-8.

Pendell, D. L., G. W. Brester, T. C. Schroeder, K. C. Dhuyvetter, and G. T. Tonsor. 2010. Animal Identification and Tracing in the United States. *American Journal of Agricultural Economics* 92(4): 927–940.

Resende-Filho, M.A., and B. L. Buhr. 2006. Economic Evidence of Willingness to Pay for the National Animal Identification in the U.S. Presented at International Association of Agricultural Economics. August, 12–18. Gold Coast, Australia.

Resende-Filho., M.A., and B.L. Buhr. 2008. A Principal-Agent Model for Evaluating the Economic Value of a Traceability System: A Case Study with Injection-Site Lesion Control in Fed Cattle. *American Journal of Agricultural Economics* 90 (4): 1091–1102. doi: 10.1111/j.1467-8276.2008.01150.x.

Scallan, E., R.M. Hoekstra, F.J. Angulo, R.V. Tauxe, M.A. Widdowson, S.L. Roy, J.L. Jones, P.M. Griffin. 2011. Foodborne Illness Acquired in the United States–Major Pathogens. *Emerging Infectious Diseases* 17(1): 7–15.

Schroeder, T., and G. Tonsor. 2012. International Cattle ID and Traceability: Competitive Implications for the U.S. *Food Policy* 37:31–40.

Seyoum, Bruk, Brian D. Adam, Candi Ge, and Eric A. Devuyst. 2013. The Value of Genetic Information in a Whole-Chain Traceability System for Beef. Paper presented at AAEA Annual Meeting in Washington, DC, August 4–6. http://purl.umn.edu/150458.

Sherman, E.L., J.D. Nkrumah, and S.S. Moore. 2010. Whole genome single nucleotide polymorphism associations with feed intake and feed efficiency in beef cattle. *Journal of Animal Science* 88(1):16–22.

Shinbaum, Samantha, Philip G. Crandall, and Corliss A. O'Bryan. 2016. Evaluating Your Obligations for Employee Training According to the Food Safety Modernization Act. *Food Control* 60: 12–17. doi:10.1016/j.foodcont.2015.07.014

Smyth, S., P.W.B. Phillips 2002. Product differentiation alternatives: identity preservation, segregation, and traceability. *AgBioForum* 5 (2):30–42. http://www.agbioforum.org/v5n2/v5n2a01-smyth.htm [accessed 08 October 2015].

Soeder, J. 1993. Anatomy of an avoidable tragedy. *Restaurant Hospitality* 77(3): 34.

Stasiewicz, M.J., H.F. Oliver, M. Wiedmann, and H. C. den Bakker. 2015. Whole-Genome Sequencing allows for improved identification of persistent Listeria monocytogenes in food-associated environments. *Applied and Environmental Microbiology* 81(17): 6024–6037. doi: 10.1128/AEM.01049-15.

Sterling, Brian, Martin Gooch, Benjamin Dent, Nicole Marenick, Alexander Miller and Gilbert Sylvia. 2015. Assessing the Value and Role of Seafood Traceability from an Entire Value-Chain Perspective. *Comprehensive Reviews in Food Science and Food Safety* 14(3):205–268.

Thakur, M. and C.R. Hurburgh. 2009. Framework for implementing traceability system in the bulk grain supply chain. *Journal of Food Engineering* 95 (4):617–626. doi:10.1016/i.jfoodeng2009.06.02 8.

Tonsor, G., and T. Schroeder. 2006. Livestock Identification. *Journal of International Food and Agribusiness Marketing* 18(4):103–118.

United States Department of Agriculture. Animal and Plant Health Inspection Service. 2015. Animal Disease Traceability. https://www.aphis.usda.gov/wps/portal/aphis/ourfocus/animalhealth?1dmy&urile=wcm%3apath%3a%2FAPHIS_Content_Library%2FSA_Our_Focus%2FSA_Animal_Health%2FSA_Traceability. [Accessed November 2, 2015].

United States Department of Agriculture. Animal and Plant Health Inspection Service. 2009. Overview Report of the Benefit-Cost Analysis of the National Animal Identification System. April (Slightly Revised). https://www.aphis.usda.gov/traceability/downloads/NAIS_overview_report.pdf

VCM International. 2014. Learning from Australia's Traceability and Meat Standards Program. http://vcm-international.com/wp-content/uploads/2014/09/Australias-Traceability-System-Case-Study-August-2014.pdf. [accessed May 13, 2016].

Weaber, R. L., and J. L. Lusk. 2010. The Economic Value of Improvements in Beef Tenderness by Genetic Marker Selection. *American Journal of Agricultural Economics* 92(5):1456-1471. doi: 10.1093/ajae/aaq062.

Wilson, K., and J.S. Brownstein. 2009. Early detection of disease outbreaks using the Internet *CMAJ* 180(8):8 829-831. doi:10.1503/cmaj.090215.

Wilson, K., B. von Tigerstrom, and C. McDougall. 2008. Protecting global health security through the International Health Regulations: requirements and challenges. *CMAJ* 179(1):44–48.

World Health Organization (WHO). 2015. http://www.who.int/foodsafety/areas_work/foodborne-diseases/en/ [accessed October 5, 2015].

Yeung, R. and W.M.S. Lee. 2012. Food Safety Concern: Incorporating Marketing Strategies into Consumer Risk Coping Framework. *British Food Journal* 114(1):40-53.

International Food and Agribusiness Management Review

The Official Journal of the International Food and Agribusiness Management Association

www.ifama.org