THE INTERNATIONAL SOCIETY OF PRECISION AGRICULTURE PRESENTS THE
13th INTERNATIONAL CONFERENCE ON
PRECISION AGRICULTURE
July 31-August 4, 2016 • St. Louis, Missouri USA

Surplus Science and a Non-Linear Model for the Development of Precision Agriculture Technology

Cushnahan M.Z, Wood B.A, Yule I.J, Wilson R

New Zealand Centre For Precision Agriculture, Massey University, Palmerston North, New Zealand

A paper from the Proceedings of the
13th International Conference on Precision Agriculture
July 31 – August 4, 2016
St. Louis, Missouri, USA

Abstract. *The advent of 'big data technologies' such as hyperspectral imaging means that Precision Agriculture (PA) developers now have access to superabundant and highly heterogeneous data. The authors explore the limitations of the classic science model in this situation and propose a new non-linear process that is not based on the premise of controlled data scarcity. The study followed a science team tasked with developing highly advanced hyperspectral techniques for a 'low tech' sector in which non-adoption by farmers is a significant risk. Hyperspectral imaging creates multi-layered, geo-referenced data early in the science process in superabundance. This data is created at high speed in near real-time and does not require expensive ground sampling. The data is extremely versatile and has the potential for many different measurements from one record. These data traits increase the likelihood of producing 'surplus science', that is, science that exceeds what was judged necessary to solve the problem as defined at project launch. The production of superabundant and highly versatile data early in the science process increases the possibility of discovering new forms of valuable knowledge (methods and solutions) during the course of an investigation. However, realizing the value of these opportunities requires a departure from the classic science model. Under data-scarcity conditions, such surplus science would be classified as undesirable 'project creep'. In response we propose an alternative process based on a non-linear, iterative approach that utilizes heterogeneous actors to refine value from hyperspectral data. The paper documents how a 'big-data' setting generates surplus science and unexpected value possibilities. We outline the challenges that science teams face if they are to realize these possibilities. These challenges include the linearity of project design and set up, which limits the ability to identify unexpected opportunities and re-organize in response. Moreover, the science team may not have either sufficient time or appropriate expertise to exploit an opportunity. In light of these findings, it is proposed that for innovation in the PA sector to make the necessary rapid advances* *both technically and in terms of adoption, changes are needed in the way research projects are funded and structured. In addition, we suggest changes to the make-up of science teams and the inclusion of a variety of end-user perspectives during the research and development process.*

Keywords. *Precision Agriculture, surplus-science, spillover, hyperspectral, value creation, big data, innovation.*

The authors are solely responsible for the content of this paper, which is not a refereed publication. Citation of this work should state that it is from the Proceedings of the 13th International Conference on Precision Agriculture. Cushnahan, M.Z., Wood, B.A., Yule, I.J., Wilson, R. (2016). Surplus Science And A Non-Linear Model For The Development Of Precision Agriculture Technology. In Proceedings of the 13th International Conference on Precision Agriculture (unpaginated, online). Monticello, IL: International Society of Precision Agriculture.

Introduction

Precision Agriculture (PA) is a rapidly evolving field of agricultural science (Mulla, 2013) and the influence of big data technologies is growing (Sonka & IFAMR, 2014). One such technology, hyperspectral sensing and imaging (HSI), proffers to radically change the role and value of data-collection for Precision Agriculture scientists. The difficulty of working with HSI data is that it not only fulfils the four V's of big data, i.e. Volume (great volume), Variety (various modalities), Velocity (rapid generation) and Value (huge value but very low density) (Chen, 2014), each georeferenced data point has many layers of data that is extremely versatile, i.e. it has the potential for many different measurements from each record.

In order to better understand the scientific process that PA researchers and developers adopt when using such data, the lead author of this study is embedded as a participant observer within the New Zealand Centre for Precision Agriculture's (NZCPA) research and development team, who are working at the cutting edge of Precision Agriculture research in action.

Pioneering to Precision

The centre is a specialist PA research group who are working on a major project; a Primary Growth Partnership called "Pioneering to Precision" funded by a co-operative fertilizer company called Ravensdown and the New Zealand Government organization, the Ministry for Primary Industries. The NZCPA researchers are tasked with developing highly advanced hyperspectral techniques using airborne hyperspectral imaging to predict nutrient concentrations in plant tissue from the air to guide and inform fertilizer decision making (Grafton & Yule, 2015; MPI, 2014).

The project aims to replace expensive, laborious and time-consuming conventional data-collection methods used for in-field plant nutrient analysis with the non-invasive, rapid collection of high volume, versatile data at a very high spectral and spatial resolution (Ravensdown, 2014; Yule, 2015). Experience from other industries suggests however, that refining value from big data technologies will be a key challenge for the agricultural sector (Huberty, 2015; Sunding, 2001; Sonka & IFAMR, 2014).

The project involved flying the hyperspectral imaging sensor AisaFENIX (Specim) over eight test farms over different seasons, with over 7,000 tissue and soil samples being sent to a laboratory for calibration. Data was then explored to see if relationships between the hyperspectral data and wet chemistry calibration samples could be found for the major pasture nutrients. Models were subsequently created from these relationships (Pullanagari, 2016).

The hyperspectral sensing and imaging creates multi-layered, geo-referenced data early in the science process in superabundance (Pullanagari, 2011; Yule, 2015). This data is created at high speed and does not require expensive and time-consuming ground sampling, although the NZCPA uses ground-truthing to calibrate its algorithms (Yule, 2015).

Early in the study we observed that the new technology was presenting the NZCPA science team with difficulties and opportunities created by what we call a new data economy. This paper articulates initial ideas and concepts generated from observations of the research and development team in action. In this paper we explore the idea that data characteristics influence the likelihood of producing what we call 'surplus science', that is, science that exceeds what was judged necessary to solve the problem as defined at project launch. We detail the effect of the arrival of superabundant and highly versatile data early in the science process on the possibility of discovering new forms of valuable knowledge (methods and solutions) during the course of an investigation.

The disruption of science

A key challenge with data-rich PA data technologies such as HSI is that the value model does not emerge fully-formed from data itself; they are simply inputs to a production process that depends on human insight (Huberty, 2015; McAfee, 2012). The Pioneering to Precision project was structured in a linear form, the Classical Model of Science (De Jong, 2010), a method that is designed to generate knowledge from limited and expensive data. The model has been widely promoted in agricultural science and produces knowledge from limited data by extrapolating a small number of results, which is then generalized for real-world application (Alrøe & Kristensen, 2002; Oliver, Bishop, & Marchant, 2013). Traditionally, the collection of agricultural research data has been a relatively slow and expensive process; data typically arrives after the experiment is judiciously planned to produce statistically significant results for a pre-determined problem. The linear, classical approach to science uses the bare minimum of samples (limited by budgets) to achieve a derived output within the scope of the problem statement; we call this approach to data collection data scarcity, see figure 1.

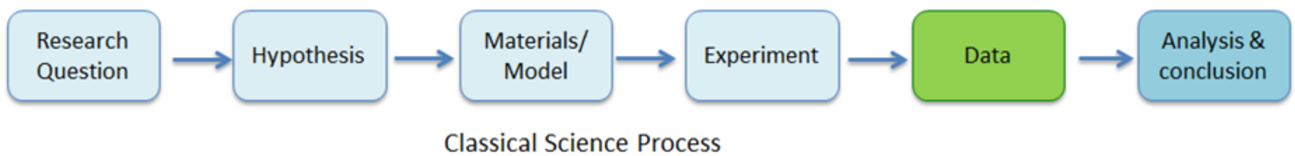

Figure 1: The Classical Science Process

PA strives to capture and respond to real-world variability and an important observation of the study was that the HSI technology generated a massive amount of highly variable and versatile data that the NZCPA science team had to explore iteratively (Sonka & IFAMR, 2014). While PA scientists are accustomed to working with variable data, they need to understand what that variability actually means in order to create the algorithms that refine the data into valuable information. We suggest that the difficulties encountered by the NZCPA scientists were exacerbated by the linear design of the Pioneering to Precision project. The project was structured and resourced as a classical science problem where scientists needed to find a way to achieve a pre-determined outcome, that is, to measure nitrogen, phosphorus, potassium and sulfur concentrations in tissue samples from the air and relate that to soil fertility in order that fertilizer application can be optimized

The NZCPA scientists embarked on an iterative data exploration process, which was particularly challenging because although there are statistical methods available in the research environment, they needed to be further developed into a more commercially operable outcome. Further they were not necessarily always sure what they were looking for in the data, as the AisaFENIX sensor had not been used for this purpose before, see Figure 2. This approach is a far cry from the linear classical science traditionally promoted for agricultural science and led us to focus on the role of data in the science-making process and to consider what a radical shift in the data economy could mean for Precision Agricultural researchers.

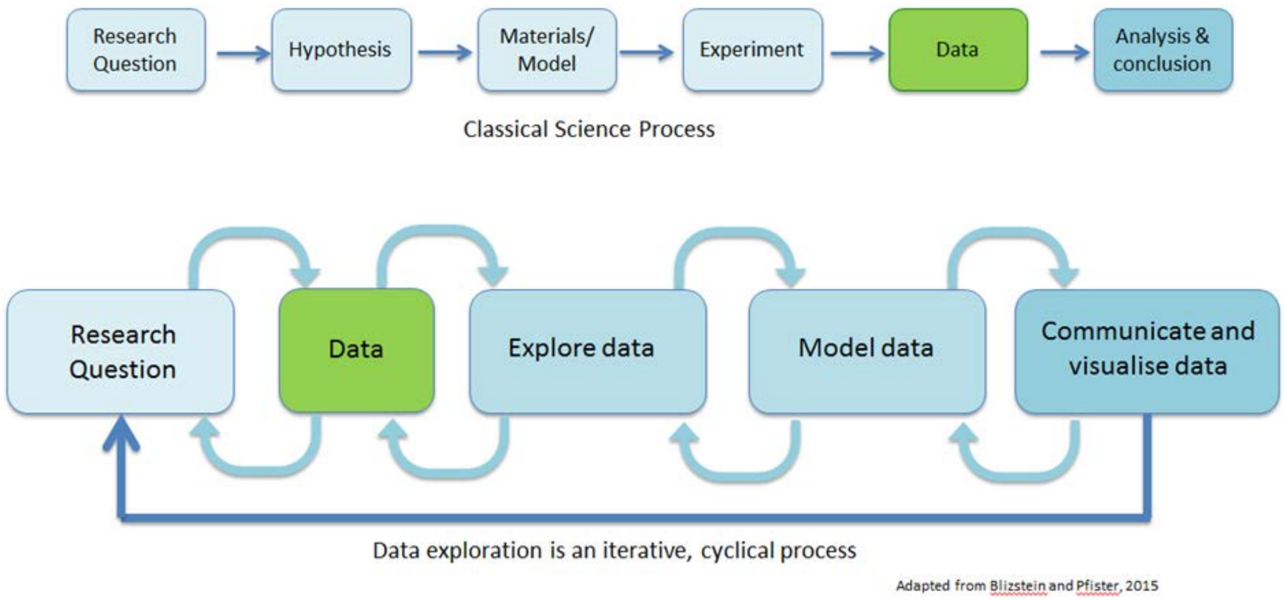

Figure 2: Disrupting the science process.

The new data economy

Some emerging Precision Agriculture technologies are radically different to their counterparts; importantly, these technologies produce data that is very different to the data that agricultural scientists are accustomed to; the data is enormous, complex and arrives early in the science process. We call this dramatic change in data characteristics, a new *'data economy'*.

Hyperspectral imaging for example, provides information at a very high spatial resolution, which can explain localized distributions of nutrients at farm or regional scale (Von Bueren & Yule, 2013; Yule et al., 2015). The data produced is relatively inexpensive and in a form that is or can be worked into an image. While this is a benefit in terms of technology transfer, observations of PA scientists working with the enormous, multi-layered data indicates that the new data economy has consequences for both the science process and the potential additional value that can be extracted from data. Understanding and addressing the complex implications of the new data economy is a key challenge for the PA scientists hoping to optimize value from these technologies.

An example of the new data economy is the nutrient testing of pastures. Soil testing for major nutrient concentrations has been a slow, relatively expensive exercise with a small number of samples taken and results are extrapolated over a larger area; data is relatively scarce and appears late in the science process, with a lot of the scientific effort being located in the planning stages. Sampling and analyzing soil samples is relatively expensive, so sample numbers are usually kept low. For example, a typical analysis of soil fertility for a large hill country farm would include taking three transects of 27 soil samples each. The analysis would be carried out on the aggregated samples from each of the three transects. The reason 27 core samples are used is that soil nutrient data from this type of source is tremendously noisy. However, any more than three transects over a 3,000ha hill country farm would be considered a generous number of samples given the cost, time and effort required to collect them. The airborne hyperspectral imaging sensor AisaFENIX has a pixel size of approximately 1 metre at 600m sampling altitude. The system can capture 30 million measurements in less than two hours, see Figure 3 (SPECIM, 2016). Importantly, not only are there substantially more geo-referenced sample points per hectare, this data is also multi-layered; so each data location yields hundreds of layers over the VIS, NIR and SWIR enabling biochemical analysis for any target (ASDInc, 2016). We call this '*data superabundance*'.

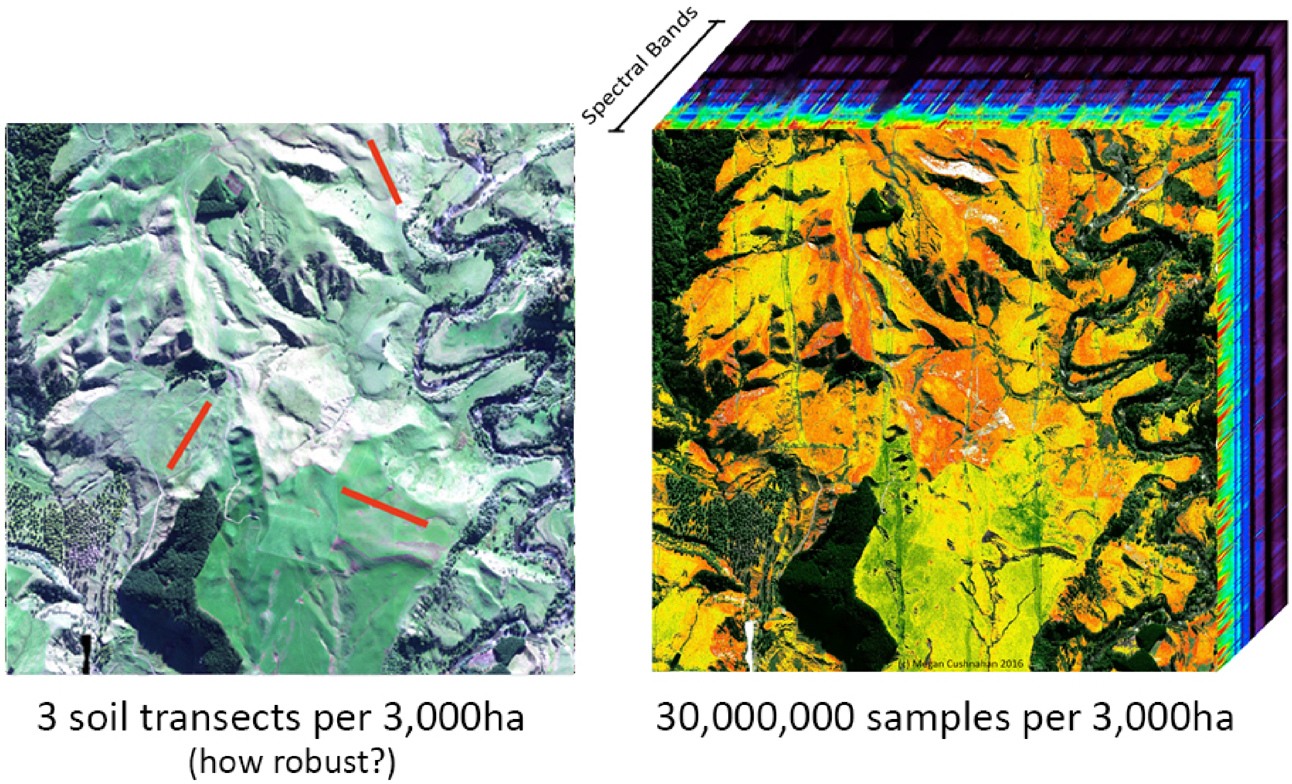

Figure 3: The new data economy for pasture nutrient testing.

While this area of research is in its infancy, we suspect that some data-rich PA tools may be more likely to generate surplus science than others. The identification of technologies that are conducive to surplus science production may help researchers adapt their science process, and project resourcing and structure to optimize the extraction of value from data.

Translating data into value

While the study shed light on the some of the risks and challenges of the new data economy, some of the huge opportunities offered by the nascent data-rich PA technologies were also revealed. Observations from the study highlighted that in the structure of the discovery process may differ with the new PA technologies. While the NZCPA scientists initially tried to follow the classical, linear structure of investigation laid out in the project plan, they were soon forced to use a cyclical, iterative approach, akin to that used in the field of data science in order to translate the superabundant data into valuable knowledge (Pfister & Blitzstein, 2013). It was discovered that the original data set may be re-used to answer new research questions encountered along the way. In terms of PA data, the spectral and spatial resolution of data, along with the volume, complexity and velocity are likely to determine the potential value of data for use in other applications.

Surplus Science

A key observation from the study was that new ideas for re-purposing the HSI data began to emerge with unusual frequency from the NZCPA team. It was identified that a single HSI data collection event not only provides the data used to predict levels of nutrients related to growth, but the data carries information that can be re-used to solve more than the pre-determined problem. While the team had some experience of data-mining, re-analyzing data from a live project to propose and answer new research questions was unchartered territory for the NZCPA team. We called the science generated using these data analyses, '*surplus science*'.

One of the new research questions that arose was whether other minerals such as copper could be modelled from the data. Historically, this opportunity may be perceived negatively as project creep and scope control would have been used to limit the science resource needed because sampling was time-consuming and expensive (Huberty, 2015). However, in the Pioneering to Precision project where data could be re-used, farmers were asked if the line of enquiry was worth pursuing early in the product-development life cycle (Katila, 2003). Responses indicated that the copper map was valuable as farmers were supplementing stock without fully understanding the copper status of their property; importantly, the information was useful from an animal health perspective. A member of the NZCPA team contacted a Massey University colleague in the field of animal health who provided valuable insight and confirmed that the potential to map copper concentrations in plant tissue was indeed of significant value. The veterinarian also proposed a third research question which meant the data was re-used yet again; this time to investigate if an antagonistic relationship between copper and molybdenum could be identified in the data, which was obtained and validated using the calibration samples from the original study. In the end, the original data was used to answer three important and valuable research questions, see figure 4. The bricolage-like approach (Yule and Wood, 2014) demonstrated that the repurposing of data could be routinely treated as a source of value, not as a risk to be avoided or minimized.

In addition, the surplus science is unlikely to have been discovered were it not for the collaboration with not only potential end-users but also scientists from other fields. If traditional experimental methods had been employed then in order to measure copper a further experiment would have had to been proposed, developed and funded separately.

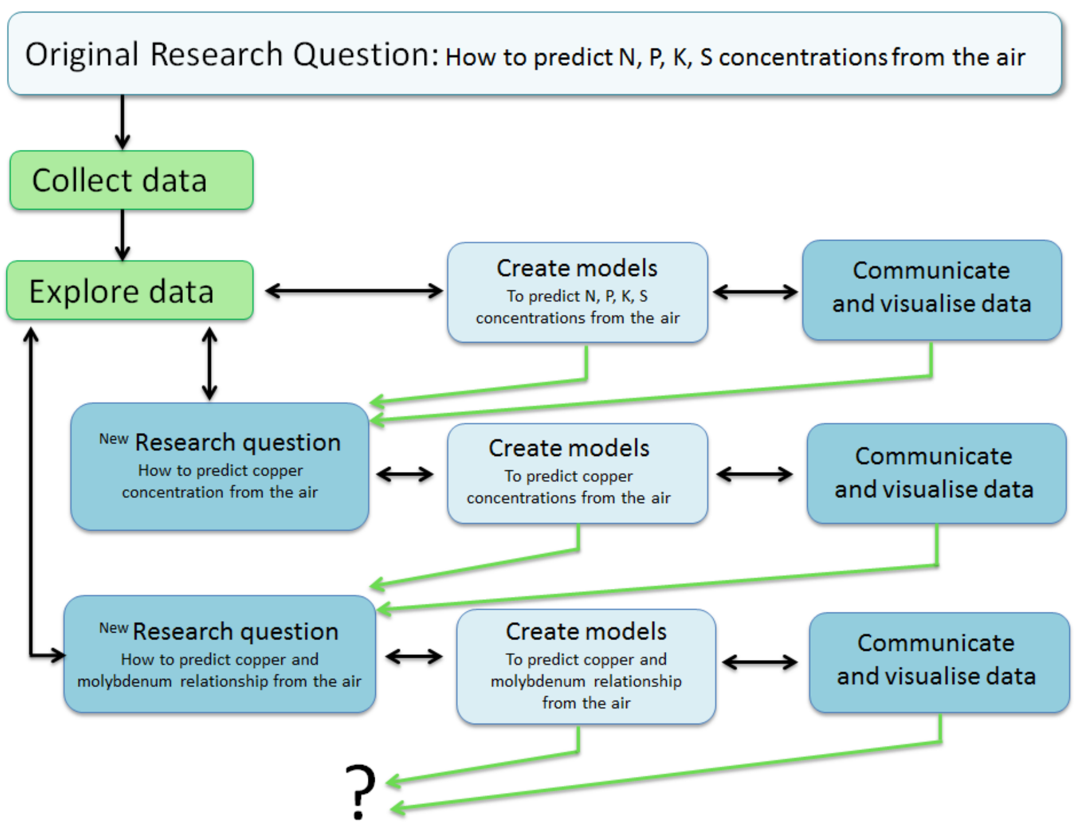

Figure 4: The iterative, cyclical science process used to generate additional value from a hyperspectral imaging project (surplus science)

Scientific Division of Labor

In addition to modifying the scientific process, observations of the NZCPA scientists also highlighted the shift in the scientific division of labor needed to translate the data into value. The traditional epistemology sees the final outcome and end of scientific activity at the early stages of the science process (Popper, 1962). A key observation of the NZCPA team was that they adapted the scientific division of labor to meet the challenge of working in the new data economy by employing more data programming, computer science and statistical analytical skills in the team than initially planned.

In addition to changes in the internal division of scientific labor, the NZCPA scientists networked beyond the end-user frame by socializing the science with a heterogeneous group from other fields to identify new opportunities and refine unexpected value from the versatile data sets. This conflicts with the linear manner in which the initial project was set up and raised questions regarding the ownership of intellectual property generated by the re-use of the original data set. This approach also conflicts with the approach that many Universities and research organisations take to fund and structure their PA research and development teams, i.e. somewhat isolated, homogeneous teams. Based on the observations of the study, we suggest that continuing to work in specialist PA teams may result in a widening gap between information and knowledge and there is a risk of missing out on value generated from surplus science.

Conclusion

A new wave of technologies producing a new data economy has the potential to disrupt research and development in Precision Agriculture. The new data economy means that Precision Agriculture Scientists are tasked with finding new ways of striking value from inexpensive, superabundant, versatile data that arrives early in the scientific process. The study of the NZCPA team indicates the new data economy may have complex and far-reaching implications for PA research and the science-making teams behind it. Understanding the data characteristics of new PA technologies will help PA researchers to adjust their science-making process and resourcing to accommodate the new data economy and to optimize the value generated from the technologies. Importantly, it is unlikely that all new PA technologies will have the same data characteristics; we need to understand how representative HSI is of emerging PA technologies and to taxonomize nascent technologies in terms of the new data economy.

Acknowledgements

The authors would like to acknowledge the New Zealand Centre for Precision Agriculture for their help and co-operation. The work on hyperspectral imaging described in this work was completed as part of the Ravensdown, Ministry of Primary Industry, (Primary Growth Partnership, PGP) sponsored project, Pioneering to Precision: Application of Fertiliser in Hill Country. The participation of the members of the NZCPA science team and participating farms is also acknowledged.

References

Alrøe, H. F., & Kristensen, E. S. (2002). Towards a systemic research methodology in agriculture: Rethinking the role of values in science. Agriculture and Human Values, 19(1), 3-23.

ASDInc. (2016). FieldSpec 4 Hi-Res NG Spectroradiometer. Retrieved from http://www.asdi.com/products/fieldspec/fieldspec-4-hi-res-ng-spectroradiometer.

Blitzstein, J. & Pfister, H., (2014). Harvard Data Science Course CS109. http://cs109.github.io/2014/.

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

De Jong, W. R., & Betti, A. (2010). The classical model of science: A millennia-old model of scientific rationality. Synthese, 174(2), 185-203.

Grafton, M., & Yule, I. (2015). The role of technology transfer to improve fertiliser use efficiency. Food Security, 7(2), 365-373.

Huberty, M. (2015). Awaiting the second big data revolution: from digital noise to value creation. Journal of Industry, Competition and Trade, 15(1), 35-47.

Katila, R., & Mang, P. Y. (2003). Exploiting technological opportunities: the timing of collaborations. Research Policy, 32(2), 317-332. doi:http://dx.doi.org/10.1016/S0048-7333(02)00102-6

McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D., & Barton, D. (2012). Big data. The management revolution. Harvard Business Review, 90(10), 61-67.

MPI. (2014). New programme set to transform hill country farms.   Retrieved from https://www.mpi.govt.nz/news-and-resources/media-releases/new-programme-set-to-transform-hill-country-farms-/.

Mulla, D. J. (2013). Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. Biosystems Engineering, 114(4), 358-371.

Oliver, M., Bishop, T., & Marchant, B. (2013). Precision agriculture for sustainability and environmental protection: Routledge.

Pfister, H., & Blitzstein, J. (2013). Course: CS109 Data Science. Retrieved from http://cs109.github.io/2014/.

Popper, K. (1962). Conjectures and refutations: The growth of scientific knowledge: Routledge.

Pullanagari, R., Kereszturi, G., & Yule, I. (2016). Mapping of macro and micro nutrients of mixed pastures using airborne AisaFENIX hyperspectral imagery. ISPRS Journal of Photogrammetry and Remote Sensing, 117, 1-10.

Pullanagari, R., Yule, I., King, W., Dalley, D., & Dynes, R. (2011). The use of optical sensors to estimate pasture quality. International Journal on Smart Sensing and Intelligent Systems, 4(1), 125-137.

Ravensdown. (2014). Transforming hill country farming returns with remote soil scanning and variable rate aerial application. Retrieved from http://www.ravensdown.co.nz/media/1879/pgp-brochure.pdf.

Sonka, S., & IFAMR, I. (2014). Big data and the ag sector: more than lots of numbers. International Food and Agribusiness Management Review, 17(1), 1.

SPECIM, S. I. L. (2016). Aisa FENIX Hyperspectral Sensor.   Retrieved from http://www.specim.fi/downloads/AisaFenix_ver1-2015.pdf.

Sunding, D. & Zilberman, D. (2001). The agricultural innovation process: research and technology adoption in a changing agricultural sector. Handbook of agricultural economics, 1, 207-261.

Von Bueren, S., & Yule, I. (2013). Multispectral aerial imaging of pasture quality and biomass using unmanned aerial vehicles (UAV). Accurate and Efficient Use of Nutrients on Farms, Occasional Report (26).

Yule, I., PullanagariI, R., Irwin, M., McVeagh, P., White, M., Kereszturi, G., & Manning, M. (2015). Mapping nutrient concentration in pasture using hyperspectral imaging. Journal of New Zealand Grasslands, 77, 47-50.

Yule, I., Wood, B. (2015). How do we transform our industry to achieve a modern vision for New Zealand Agriculture? , In: Moving farm systems to improved attenuation. (Eds L.D. Currie and L.L Burkitt).

Yule, I. J., Wood, B. (2014). Precision Agriculture as Bricolage: Understanding the Site Specific Farmer. Paper presented at the 12th International Conference on Precision Agriculture, Sacramento, California USA.