Methodological News
A Quarterly Information Bulletin
Australian Bureau of Statistics
ABS Methodology and Data Management Division
December 2014

Articles

Realising the Potential of Satellite Imagery to Estimate Official Crop Area Statistics

The ABS is currently investigating the potential of satellite imagery data to aid in the production of agricultural statistics, such as land use, crop type and crop yield. Satellite data is seen by the ABS as a useful data confrontation source, and in time, may supplement information directly collected in the agricultural program. Therefore scoping out and demonstrating the ability of this big data source to produce comparable and/or more sophisticated agricultural statistics is a key driver behind current research efforts. Success in this space may see significant reductions in provider burden and improve the timeliness of available data and thus extend the array of stakeholder requirements met or exceeded.

The Analytical Services Branch (ASB) has been working with the subject matter and technology areas in ABS to scope out future directions for the project. This has included consideration of external stakeholder needs and any constraints that may be introduced from a technology infrastructure perspective given the magnitude of big data in scope. This has resulted in the formation of three primary future directions:

Establishing partnerships: Nurturing relationships with experts in the field is essential in fast-tracking our own project work. It is also viewed as an ideal opportunity to build the ABS reputation via leading big data discussions and sharing of experiences. ASB has contributed to an international collaboration on big data through the UNECE High Level Group for the Modernisation of Statistical Production and Services. Ongoing engagement has also been established with CSIRO, Australian Bureau of Agricultural and Resource Economics and Sciences (ABARES) and Geoscience Australia, given the wealth of detailed satellite imagery data that can be remotely accessed through their collaborative computing installation - the National Computational Infrastructure (NCI, see http://nci.org.au).

Method development: Gaining an understanding of the sophisticated methods available and the assumptions underpinning them is essential in motivating future research directions. This enables ASB to make an informed assessment of which methods can be appropriately adopted to assist in addressing key research questions and thus meet the needs of stakeholders. Specific methods include state space modelling approaches for leveraging off spatial and temporal trends in the data, machine learning algorithms and parametric modelling approaches. These all seek to classify crop types based on surface spectral reflectance information collected by satellite sensors.

Acquisition of data sources: ABS requires additional data sources in order to confidently test and validate statistical methods and determine the extent to which these may be applicable in practice. There has already been some progress in this space with ground truth data acquired for some regions which enables informative statistical testing. Geoscience Australia has also provided the ABS with satellite surface-corrected reflectance data, which is one of the richest data sources of this nature in Australia. Ultimately ASB views this initial extract of

data as a step towards signing up to a more sustainable position of analysing this high quality satellite imagery directly through the NCI. Both sources will see the richness of training datasets improve dramatically in the near future, which will allow for better method assessment.

A Methodology Advisory Committee (MAC) paper was presented in June 2014 appraising potential statistical approaches to extract relevant information from satellite imagery data - see Marley, Elazar & Traeger (2014). It is planned that further method development in this area will see an additional MAC paper focussed on applying state space modelling (SSM) approaches, to be delivered in June 2015. This paper will include results from evaluating the use of SSMs for classification of agricultural crops using the aforementioned test datasets.

References

Marley, J., Elazar, D. & Traeger, K. (2014) 'Methodological Approaches for Utilising Satellite Imagery to Estimate Official Crop Area Statistics', cat. no. 1352.0.55.144, Australian Bureau of Statistics, Canberra (not yet released).

Tam, S. & Clarke, F. (2014) 'Big Data, Official Statistics and Some Initiatives by the Australian Bureau of Statistics', *International Statistical Review* (provisionally accepted).

Further Information

For more information, please contact Jennifer Marley (02 6252 5760, jennifer.marley@abs.gov.au) or Ryan Defina (02 6252 7779, ryan.defina@abs.gov.au).

Methodology Architecture – A Roadmap for New Methodological Directions in the ABS

To maintain its strong brand as a central statistical agency into the 21st century, the ABS is transforming the way it acquires, collates, uses, reuses and disseminates statistical information. To support this transformation, innovative, industrialised and contemporised statistical methods and tools will be required. Methodology Architecture provides a roadmap for systematically assessing and developing these 21st century statistical methods and tools covering the full spectrum of the statistical production cycle.

The ABS transformation vision comprises a products vision and a process vision. The ABS' products vision is a 21st century National Statistical Organisation (NSO) which has the ability and agility to combine ABS data with other strategic sources of data, including administrative data, transactional data, Big Data, and "organic" data, to produce more timely and relevant official statistics. As well, the process vision sees the ABS fully embracing industrialisation and standardisation in its business model in the production of statistics.

The Methodology Architecture vision for the ABS is "to provide a set of methods that underpins the products and process visions of the ABS transformation program." Fundamental to achieving this vision are five key elements in developing the future inventory of statistical methods and tools: innovate, industrialise, contemporise, build capability, and build support.

Innovate: New methods and tools are required to address the new and emerging challenges from the ABS products vision. An example of this is in the linking of data to create fused unit record files or new statistics, and their processing, analysis and dissemination at the micro and macro data level. The goals are to improve relevance, production cycle time, organisational capability and meet legislative requirements to protect the confidentiality of personal or business information.

Industrialise: In the post transformed world, we see significant improvement in the production cycle time through re-use of data available within and without the ABS, as well as through re-use of statistical processes, statistical methods and tools. For this to occur, the post transformation methods and tools need to support plug and play, standardisation and corporatisation, being connectible and metadata driven, and being user-driven.

Contemporise: The survey methods used by the ABS are predominately design based (often model-assisted) and in time series analysis predominately filter based (X11, X12ARIMA) methods are used. There is an aspiration for more use of statistical models to guide our choices of estimators in our methods. In addition, responsive design methods are starting to be used to improve the cost-efficiency of survey collections and reduce survey errors.

Build Capability: The ABS framework for building methodology capability is based on attracting the best budding methodologists from university graduates, offering comprehensive training to graduates, offering interesting, challenging and varying work as well as providing opportunities to attend conferences in order to retain staff, peer reviewing key methodological work and collaborating with several universities for research, technical advice and teaching short courses.

Build Support: It is as important to build support from the senior managers of the ABS as from methodologists for the methodology architecture. We found it very useful to establish a small team of champions who are enthused about the new methods, expose them to the new techniques, and encourage them to promote these new ideas to their colleagues.

To be successful, the Methodology Architecture must address all 5 key elements satisfactorily. Each of the elements poses its own challenges, and we will work with academics, experts in the field, other NSOs as well as ABS stakeholders, including methodologists, to address them.

A more detailed version of this article is to appear in the next edition of the *Statistical Journal of the IAOS*.

Further Information

For more information, please contact Siu-Ming Tam (02 6252 7160, siu-ming.tam@abs.gov.au).

Respondent Engagement in Multi-Modal Household Surveys

In December 2012 the ABS introduced a self-complete mode to one of its household surveys, the Monthly Population Survey (MPS) through an online version of the survey. As the web form mode offers increased efficiency of data collection it is now aimed to be made the primary mode of data collection in the MPS. To enable this, Data Collection Methodology (DCM) adapted a variety of techniques and developed a staged correspondence model in order to engage effectively with MPS respondents through written correspondence. This approach has been effective in increasing the proportion of respondents in choosing to opt for web form over interview modes.

The staged correspondence model that has been introduced involves:

1. Pre-approach: provides an engaging advance warning of the online survey request, informing respondents of the value of the survey.
2. The approach: provides a clear request and easy-to-follow instructions to fulfil this request in order to elicit a positive response from respondents.
3. A reminder: provides another opportunity to convert households to the web form by acting as a prompt of the previous request to elicit a positive response.

Within each piece of correspondence, the following techniques were used:

1. Improving the visual design of the correspondence packaging mailed to respondents to make the purpose of the mailing clearer from the outside in order to encourage higher opening rates.
2. Implementing visual design principles *within* the package, including a timeline and better targeted graphics to improve the clarity of communications to ensure that key messages are clearly explained and can be followed by respondents with ease.
3. Personalising the communication by seeking the respondent's individual contribution to the survey, by using personal language to ensure that the respondent feels like they are dealing with a human rather than a machine and by framing the broader context in terms that the respondent can personally relate to.
4. Utilising Cialdini's (1983) principles of influence to encourage overall participation and specifically to respond online. For example, social norms have been used by presenting participation in the survey as the norm.
5. Presenting communications in plain language so that the message presented is easily understood by respondents even if they only have time to skim read the correspondence. Although this is a guiding principle of good communication, it is particularly important when the primary method of a request is written.

These introductions saw households registering their contact details so that they would be eligible to complete the survey online increase from 23.74% to 30.73%. This shows that the new approach has caused registrations to increase by almost 30%.

References

Cialdini, R. B. (1983). *Influence: the psychology of persuasion*, The Business Library, Melbourne.

Further Information

For more information, please contact Alana Cook (02 6252 7071, alana.cook@abs.gov.au).

Sample Design for Additional Selections from Growth Hubs in the Business Characteristics Survey

The ABS received user funding from the Department of Industry (DoI) to survey additional businesses in the 2013-14 Business Characteristics Survey (BCS). The aim was to produce a set of benchmark indicators for five key sectors identified by the DoI as critical to the ongoing strength and growth of the Australian economy, known as "growth hubs". The five growth hubs are:

1. advanced manufacturing
2. mining equipment, technology and services
3. oil, gas and energy resources
4. food products and agribusiness
5. medical technologies and pharmaceuticals.

More information about DoI's Industry Growth Centre Initiative can be found on this webpage: http://www.industry.gov.au/industry/Pages/Industry-Growth-Centres.aspx#header

The additional businesses selected for the growth hubs sample will be asked a subset of the questions given to businesses in the BCS and the Business Longitudinal Database (BLD).

The ABS worked with DoI to identify the groups of Australia and New Zealand Standard Industrial Classification (ANZSIC) classes that best matched the five growth hubs. The growth hubs were not mutually exclusive. For example, some ANZSIC classes were included in both the "Advanced Manufacturing" and "Medical Technologies and Pharmaceuticals" growth hubs.

The ABS Business Survey Methodology (BSM) area then developed a sample design to meet standard error constraints at various levels, such as growth hub by ANZSIC division and growth hub by reported employment class. The key variable of interest was innovation rate (for innovations in goods and services, operations, organisation, or marketing). A single sample allocation process was implemented for all the growth hubs at once, rather than running a separate allocation for each of the growth hubs. This approach allowed the allocation process to take into account strata that were in scope of more than one growth hub.

The growth hub selections were designed to minimise respondent burden. BSM selected as much of the growth hubs sample as possible from the samples that had already been selected for the 2013-14 BCS and the most recent three panels from the BLD. The remaining sample was selected to minimise the overlap with current and future BCS and BLD samples and other major economic collections. There is funding for repeating the additional sampling from growth hubs in a future year. However, at this stage the timing of the next growth hubs collection is uncertain. The selections were designed so that if the collection is repeated in the following year, there will be some sample in

common with the current selections. However, if the collection is repeated in three years' time, it may not be possible to achieve much common sample.

Further Information

For more information, please contact Victoria Leaver (02 6252 5445, victoria.leaver@abs.gov.au) or Lyndon Ang (02 6252 5279, lyndon.ang@abs.gov.au).

The Quality Challenge Presented by Big Data

The ABS is involved in an international task team to investigate and address quality issues in using big data for official statistics purposes.

The amount of data available in digital form is increasing in volume, variety and speed of access, a phenomenon known as the emergence of 'Big Data'. This presents both opportunities and risks for National Statistical Offices (NSOs): opportunities created by the new data sources for new products and analyses, as well as the risk of being 'left behind'. An additional risk is that the quality of the new sources of information and the statistical products derived from them might not meet the quality standards required of official statistics.

The High-Level Group for the Modernisation of Statistical Production and Services (HLG; formed by the Conference of European Statisticians Bureau) identified Big Data as a key challenge for official statistics. As a consequence, a project to investigate Big Data for official statistics, *The Role of Big Data in the Modernisation of Statistical Production,* was undertaken with the United Nations Economic Commission for Europe (UNECE) undertaking the project management and secretariat functions. The project began in July 2014.

The ABS has been playing a key role across a number of activities of this international collaborative project that includes multiple NSO representatives. One area where we have been particularly active is in the project's Big Data Quality Task Team, with the ABS taking the leadership of this team.

The goals of the team were to assess existing statistical data frameworks as to their suitability for Big Data and then produce a framework that explicitly addressed Big Data if required, test that framework with test datasets produced by other teams within the project, and propose quality indicators that might be useful in a Big Data environment.

This work is almost complete and a summary of the team's work so far was presented to the HLG meeting in Geneva in November. The final draft paper will be circulated to members of the ABS big data forum for consultation.

The work represents the first steps in formulating a systematic approach to data quality with Big Data for National Statistical Offices.

Further Information

For more information, please contact David Dufty (02 6252 6187, david.dufty@abs.gov.au).

Moving Trading Day Adjustment

It has long been understood that trading day variations within months can represent a very significant proportion of the movement in many time series, and that being entirely calendar related, these effects should be removed as part of our standard seasonal adjustment procedure.

One of the major difficulties in accounting for trading day effects is that they are not static and vary considerably through time. For example, as Sunday trading has become more popular over the span of the last few decades, a portion of trading activity has shifted to Sunday from the other days of the week. An improvement to the standard X11 static trading day regression was made by the Time Series Analysis section and the technique has been implemented in the ABS's seasonal adjustment package SEASABS since 1998. SEASABS allows the estimated daily weights to vary over time by performing multiple trading day regressions on sub-spans of the data. The daily weights are smoothed so that appropriate moving daily weight estimates are available for every year in the data span. This leads to substantial improvements in our seasonally adjusted estimates over those estimated using a static trading day adjustment.

The ABS treatment of Moving Trading Day is unique among National Statistical Organisations (NSOs), compared with the current standard treatment of estimating Trading Day variation as a constant (simple average) effect over the analysed span of a time series. Other NSOs have recently expressed interest in the ABS approach. In response to this interest, the Time Series Analysis Section is preparing a research paper on this topic, based on earlier unpublished work. Additionally, we are currently working on further improvements that will keep the ABS at the forefront of research in this area.

The research paper *Trading Day Adjustments in SEASABS* will be released in early 2015.

Further Information

For more information, please contact Lujuan Chen (02 6252 6342, lujuan.chen@abs.gov.au) or Jonathan Campbell (08 9360 5383, jonathan.campbell@abs.gov.au).

How to Contact Us and Email Subscriber List

Methodological News features articles and developments in relation to methodology work done within the ABS Methodology and Data Management Division. By its nature, the work of the Division brings it into contact with virtually every other area of the ABS. Because of this, the newsletter is a way of letting all areas of the ABS know of some of the issues we are working on and help information flow. We hope the Methodological Newsletter is useful and we welcome comments.

If you would like to be added to or removed from our electronic mailing list, please contact:

Peter M Byron
Methodology & Data Management Division
Australian Bureau of Statistics
Locked Bag No. 10
BELCONNEN ACT 2617

Tel: (02) 6252 6804
Email: methodology@abs.gov.au

The ABS Privacy Policy outlines how the ABS will handle any personal information that you provide to us.