Collecting Geo-Referenced Data in Farm Surveys

Philip Kokic[1], Kenton Lawson[2], Alistair Davidson[3], Lisa Elliston[2]
Australian National University[1]
Australian Bureau of Agriculture and Resource Economics[2]
Productivity Commission[3] *

Abstract

This paper describes a new capability recently developed by the Australian Bureau of Agriculture and Resource Economic (ABARE) for the collection of paddock level geo-referenced and land use data in its main farm surveys. A participatory approach to data capture aided by interactive computer technology is used. There are many potential benefits of this capability, but two which are described in this paper are an improvement in the respondents' involvement in the interview process and an improvement in economic modeling through the incorporation of spatial covariates.

Keywords: Spatial data, Regression, Linking data

1. Introduction

This paper provides an overview of the expanding survey capability of ABARE for collecting geo-referenced paddock level data. This capability is primarily about integrating science and economics at a finer spatial resolution in order to deliver cost effective support to a broader range of policy issues. Substantial changes have been made to the way in which ABARE farm survey collections are undertaken and the way in which the survey data can be analyzed in order to address these information shortfalls. In particular, data can now be collected within a GIS framework. Farm boundaries, primary management areas (that is crop rotation, rangeland and nonagricultural areas) as well as individual survey year crops are identified by the farmer. They are then mapped by the survey officer as part of the interview to collect data on the physical and financial performance of surveyed farms.

These techniques have recently been integrated with ABARE's broadacre and dairy surveys for the first time (Davidson, Elliston, Kokic and Lawson 2005). In 2005, ABARE surveyed 1875 broadacre (that is, farms whose main production is beef, sheep, wool and/or cereal crops) and dairy farms with a combined operating area of around 60 million hectares across Australia. With roughly half of these properties mapped to some extent, the potential for acquiring large volumes of spatial data at a sub-farm level is readily apparent.

Some of the potential benefits of collecting geo-referenced data at paddock level from surveyed farms are as follows:

- Development of a greater capacity to ground-truth models based on satellite data that are increasingly used for natural resource management.
- Improved estimation of greenhouse metrics such as carbon sequestration on agricultural land and emissions from agriculture.
- An increased capacity to generate small area estimates that reflect the heterogeneity within and across landscapes. These estimates have a wide range of uses in government.
- An ability to undertake economic analysis of changes in land management practices based on the reliability of water supply and rainfall.
- Improved methodologies for providing higher quality and more timely production forecasts for major Australian agricultural commodities through the capacity to analyse spectral signatures of crops and pastures using satellite imagery.
- A better understanding of the economic impacts of pest and disease incursions on farms using finer resolution spatial data to improve the evaluation of post-incursion management options.
- A reduction in the number of variables that need to be collected in surveys, resulting in reduced response burden.

* This research was conducted at ABARE and the content of this article does not necessarily reflect the views of the Productivity Commission.

In addition, the use of effective computer capture techniques during the interviews has changed the collection of data from the traditional question-answer approach to an interactive session in which digital maps of the property and the structure and geography of land use on the farm is obtained. The new approach has improved farmers' involvement and responses to the survey despite an increase in 'traditional' response burden. Consequently, it is argued that response burden needs to be considered more holistically along with any positive benefits resulting from the participatory collection method.

In this paper a description is given of the farm mapping tool, survey approach and data analysis methodologies that have been developed so that farm level financial and physical data can be linked with paddock level natural resource data. Also, a simple application is presented to highlight the potential benefits of the linked data for the modeling of farm land values.

2. About the Spatial Mapping Tool

ABARE developed a spatial mapping tool known as SGP (survey geo-referencing project) in order to collect geo-referenced and spatial data during face-to-face interviews on farms (figure 1). The spatial data collected with SGP enable the integration of farm-level financial and physical data with sub-farm spatial data such as land capability and vegetation cover.

SGP was developed in Microsoft Visual C++ 6.0 using MFC, ADO, Shapelib C library v1.2 and Earth Resource Mapping's ECW JPEG2000 SDK v3.1. When working in geographic coordinates the areas of digitized polygons are computed using an Albers equal area projection. Digitized polygons may be queried and edited and self-intersections are not allowed. Polygon and point attributes are stored in a Microsoft Access database as well as in shape files. Printing and export to BMP are also supported. Standard GIS navigation tools are provided, as illustrated in the navigation toolbar in figure 1. Digitized polygons can be edited in the field using purpose built tools.

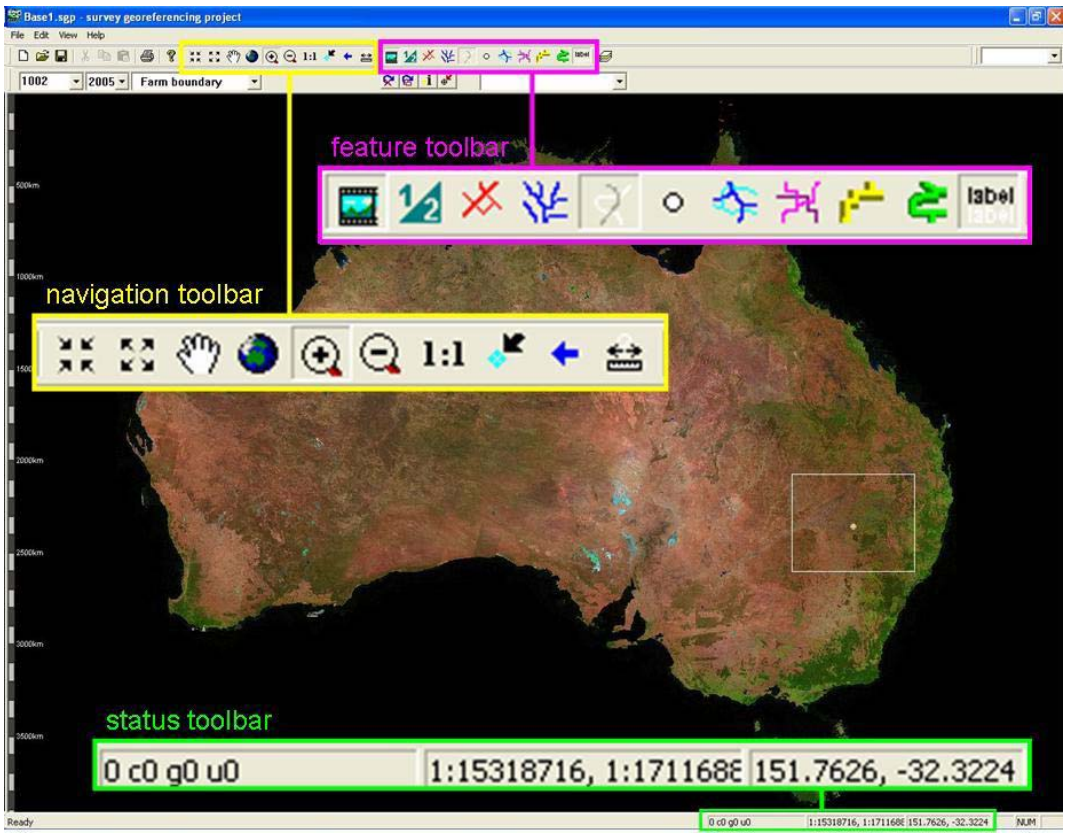


Figure 1: Geo-reference mapping tool toolbars

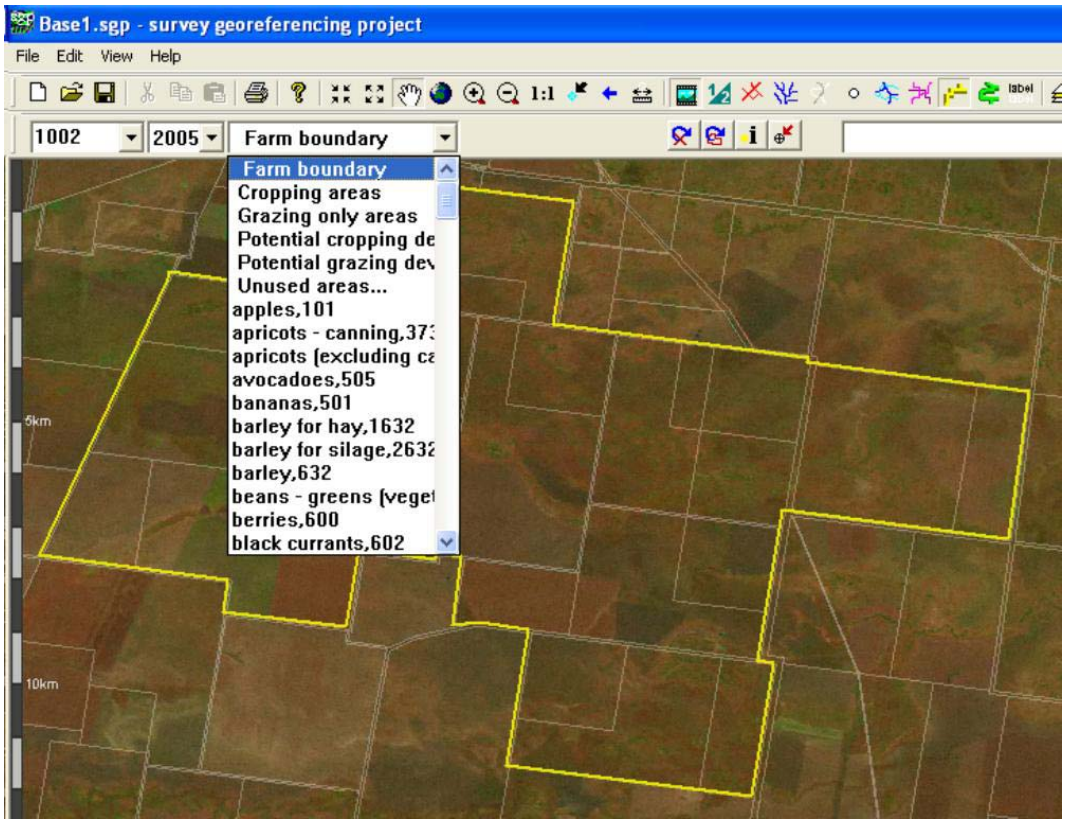


Figure 2: Defining farm areas to be mapped

The background image that has been adopted for farm mapping is a landsat satellite image mosaic of Australia, usually provided by the Australian Greenhouse Office, overlaid with cadastral boundaries, as illustrated in figure 2. Ideally, a landsat image is selected that best captures seasonal activities during the survey year to facilitate farmers' identification of management and cropping areas although other recent images have been used.

Other feature layers that are used to assist in locating farms and identifying areas to be mapped are: roads, rivers, state boundaries and localities derived from Geoscience Australia's Topo250k data sets, river basins, Australian Bureau of Statistics (ABS) Statistical Local Areas, and ABARE broadacre and dairy survey regions, as depicted in the feature toolbar in figure 1. Some feature labeling and search facilities are also supported.

The status toolbar shown in figure 1 contains various summary statistics of the current mapping operation including polygon areas, horizontal and vertical scales and the latitude and longitude of the cursor's position

Farm mapping is undertaken using a point and click technique on vertices to create polygons. Polygons with holes can also be digitized. The polygon data is stored in a database along with a unique farm number and collection year to enable these spatial data to be merged with farm financial and physical data. Auxiliary non-spatial data, such as management area type, crop yield or type of vegetation cover, are also collected concurrently with the farm areas being mapped to enable further interpretation of the spatial data.

3. Collecting Geo-Referenced Data on Farms

3.1 Mapping Farm Management Areas

The majority of ABARE farm surveys are undertaken in person in a farm office or around a kitchen table. Farm spatial data capture begins by digitizing the 'farm boundary' polygon to define the farm operating area (figure 2). The farm operating area includes freehold land, areas of private and crown land leased and areas share-farmed but excludes areas leased out.

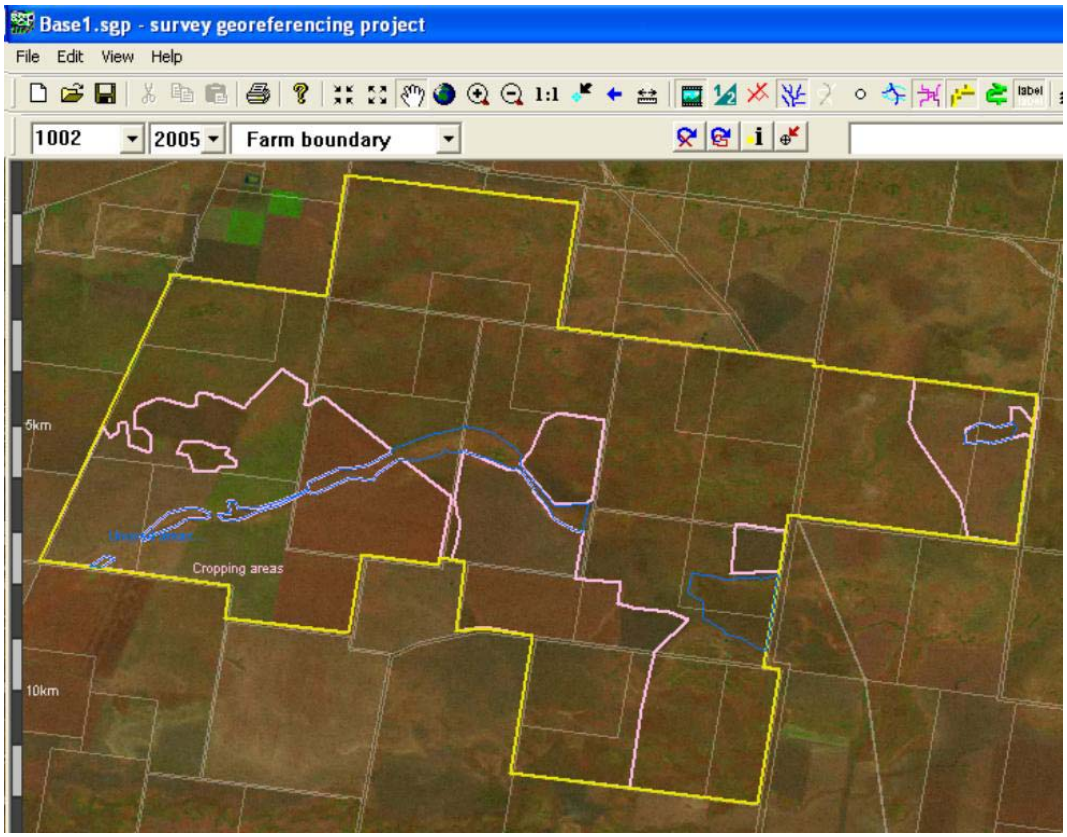


Figure 3: Mapping primary management areas

The farm boundary polygon is defined by repeatedly clicking to record all the places where the boundary line changes direction. The current polygon area is shown in the status bar and survey officers check that the calculated area is within, say, 20 per cent of the farmer's expectation. Mapping accuracy is increased after the data are cleaned up in the office - particularly where boundaries follow complicated paths such as stream lines.

Digitizing other farm polygons is the same as for the farm boundary except that the appropriate description is selected from the list box. The primary management areas are usually digitized next and these include crop rotation areas (areas that can be cropped or grazed, including fallow areas), grazing only areas and areas that are not directly used or are unsuitable for agricultural production (including airstrips, dams, heavily timbered areas or steeply sloping country). In the demonstration property (figure 3), the crop rotation and unused areas are in pink and blue respectively. In this case, the balance of the property is grazing only areas. It is generally more accurate to omit this major management type during surveys as it can be generated in the office by subtracting the other management areas from the farm boundary.

Survey year crop areas can also be digitized and where farmers keep individual paddock records, the production can also be recorded against the area (figure 4). This facility has been included to enable the creation of paddock level productivity models.

3.2 Cleaning Geo-Referenced Survey Data

Cleaning of the geo-referenced farm survey data is undertaken in the office after collection to improve accuracy and maintain quality. It involves the use of a combination of manual and automated techniques.

The data shape files are edited in ESRI ArcMap™ 9.1. Farm boundary areas are snapped to (i.e. adjusted so as to confirm with) geometric cadastral boundaries and traced along water courses and other irregular features. The two mapped primary management areas are visually edited against the background satellite image and then snapped to farm boundaries at given tolerance

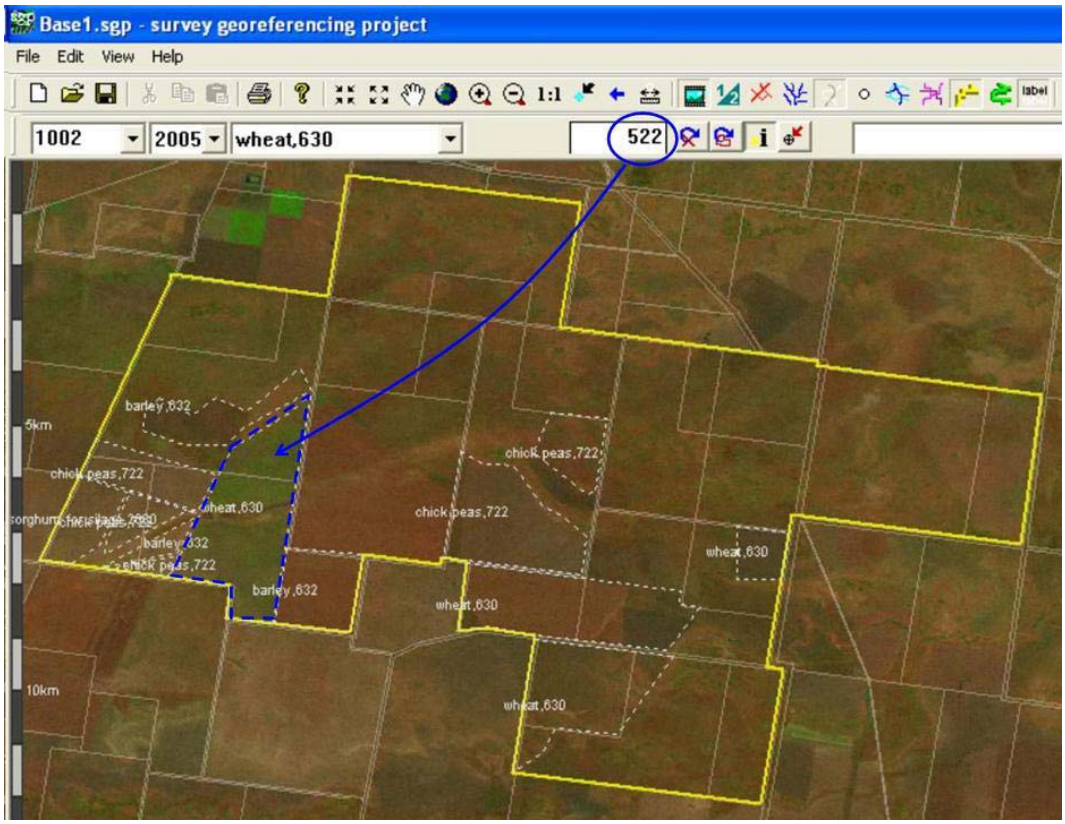


Figure 4: Mapping survey year crop areas

levels. The residual unmapped management area is clipped out of the total farm area to ensure that all primary management areas are mutually exclusive and all inclusive. Survey year crop areas are snapped to crop rotation areas and visually edited to match the underlying image where possible. Existing crop rotation areas are clipped out of potential grazing and crop development areas to ensure logical consistency of land development.

The areas of survey polygons are also cross checked against farm management areas recorded in the financial and physical survey.

3.3 Response Burden

With the large questionnaire typical of ABARE farm surveys response burden has become a significant issue.

Using effective computer capture techniques during the interviews, the method of data collection has changed from the traditional question-answer approach to an interactive session in which digital maps of the property and the structure and geography of land use on the farm is obtained. The new approach has improved farmers' responses because it facilitates communication with the farmers in a way they view their property – visually. The overwhelming experience has been that the farmers find this part of the survey most interesting. Feedback received from interviewers suggests that this has lead to a greater involvement of the farmers in the interview process and potentially an improvement in the quality of data collected.

Cadastral farm boundary data has enabled ABARE to link its farms with crop yield and crop area models, and thus effectively provide continuously updated crop information throughout the growing season. For example, the shire scale wheat and sorghum yield models of the Queensland Department of Primary Industries provide in season distributional forecasts of crop yields across the crop producing regions of Australia (Potgieter, Hammer and Butler 2002). This model is based on a soil water balance model, and depends on soil data, climate station data, and is calibrated to ABS agricultural census data. More

recently, a harmonic analysis approach using MODIS Enhanced Vegetation Index satellite images plus ground truth data was used to determine crop types and areas sown (Potgieter, Apan, Dunn and Hammer 2007). The potential for reducing response burden can clearly be seen from these last two examples. In the medium term biophysical and statistical modeling have the strong possibility of substantially reducing the number of variables that need to be collected.

Consequently, it is proposed that response burden needs to be considered more holistically along with any positive response benefits resulting from geo-referencing sample farms and the participatory collection method. One way this could potentially be achieved is by accounting for the additional information each question may make available for analysis and by weighting the degree of relevance respondents may view each question. For example, questions used for geo-referencing paddock boundaries are generally viewed as highly relevant by respondents and they also enable the addition of a large amount of information for analysis through linkage with external spatial data sources, and hence such questions should substantially reduce any measure of response burden.

3.4 Sources of Spatial Data

ABARE works with both state and other federal government agencies to obtain spatial data layers. ABARE models use spatially derived covariates based on both shapefiles and raster datasets. Area data are generated by intersecting farm survey polygons with other shapefile data layers such as vegetation density and land capability. Raster or grid-point data such as climate or moisture indexes are averaged over farm survey polygons.

4. ABARE's Farm Surveys

Collection of geo-referenced data can now be fully integrated with ABARE's annual farm surveys: the Australian Agricultural and Grazing Industries Survey (AAGIS) and the Australian Dairy Industries Survey (ADIS). These surveys are used to collect detailed financial, physical and socioeconomic information from farm businesses across Australia (ABARE 2004).

The data collected via AAGIS, ABARE's main farm survey platform, are comprehensive, covering all agricultural and nonagricultural production; costs and receipts and inputs and outputs. These data correspond to the area actually operated by the farmer, with adjustments made for areas leased or share-farmed.

The main method of collection for the broadacre and dairy industry surveys is face to face interviews with farmers with the mapping and other data collected electronically on laptop computers. Although face to face interviews are resource intensive, this approach also assists in maintaining a high level of quality control over the data that are collected.

4.1 Sample Selection

ABARE surveys are designed and samples selected on the basis of a framework drawn from the Business Register maintained by the ABS to conduct its agricultural census. This framework includes agricultural establishments in each Statistical Local Area classified by size and major industry. Sampling from the ABS frame ensures that statistically reliable estimates can be made.

The surveys cover agricultural establishments with an estimated value of agricultural operations of $22 500 or more. Industry definitions are based on the Australian and New Zealand Standard Industry Classification (ANZSIC) (ABS 2006). This classification is in line with an international standard that is applied comprehensively across Australian industries, permitting comparisons between industries, both within Australia and internationally.

4.2 Sample Weighting

The population of broadacre farms in a survey region is typically stratified and the sample selected in varying proportions from the strata to improve the precision of survey estimates. Since the mix of enterprises varies from farm to farm, a consequence of this approach is that sample weights must differ from farm to farm.

Sample weights are calculated using a model based weighting procedure (Bardsley. and Chambers, 1984) so that population estimates of numbers of farms, areas of crops and numbers of livestock in various geographic regions and industries correspond as closely as possible to known totals supplied by the ABS. The same procedure was used to further adjust the weights so that estimates of total agricultural areas of different land capability types would conform to known totals. Once the farm paddock boundaries have been mapped other spatial layers can potentially be linked and used as spatial covariates for weighting of the survey data.

5. Application of the Geo-Referenced Data to Modeling Broadacre Land Values

Broadacre agriculture is large scale cropping and grazing activities relying on extensive areas of land with a minimum of labor. From an agricultural perspective, the productivity of such land can be significantly dependent on access to other natural resource factors such as water and vegetation. Where such access is regulated beyond what farmers consider to be optimal, land productivity may decrease. In the long run, changes in access to natural resources are ultimately expressed in changes to agricultural land values. In broadacre areas where the value of farm land is largely dependent upon its agricultural income earning potential, land value models are useful for estimating the long run impacts of such policy changes.

In order to investigate how vegetation cover is related to farm values, in 2005 ABARE surveyed about 250 broadacre farmers situated in central and western New South Wales. The study region covered an area of approximately 400,000 square kilometers (figure 5). This region is characterized by cattle, wool, prime lamb and extensive dryland cropping operations on farms with diverse levels of vegetation. It also contains farms that are under multiple adjustment pressures, areas with significant development potential and properties affected by native vegetation regulations.

Hedonic regression techniques have long been used to evaluate factors that influence agricultural land values (Palmquist 1989; Palmquist and Danielson 1989). Accordingly, a hedonic model has been used to explain broadacre land values in the study region in terms of land characteristics that vary between farms, as shown in the equation below:

$$\log(V_j) = \beta_0 + \sum_i \beta_i Z_{ij} + \varepsilon_j, \quad (1)$$

where V_j is the value of broadacre agricultural land (in dollars per hectare) less depreciated values of the operator's house and other capital improvements, Z_{ij} is the set of explanatory covariates and ε_j is a random error term with mean zero and constant variance. The β_i coefficients are the marginal values per hectare of

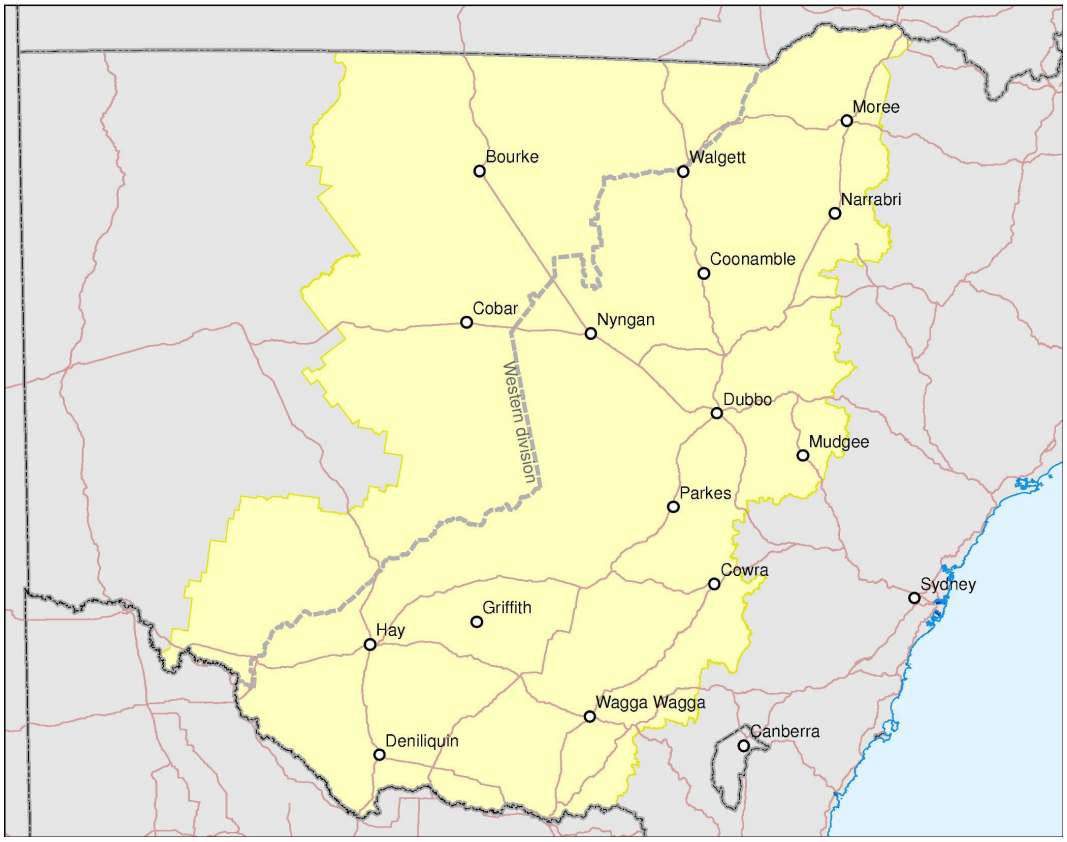


Figure 5. Study region in New South Wales

respective land characteristics. The value of farm land V_j is self-assessed by farmers participating in ABARE's AAGIS survey and is highly correlated with independent valuations. Potential covariates that were considered are described below:

- **Climate** is measured by the forest productivity index; a relative climate index of long term plant productivity sourced from the Australian Greenhouse Office. The forest productivity index model is based on the relationship between the amount of photosynthetically active radiation absorbed by plant canopies and the various productivity modifiers that affect plant growth including temperature and soil water content (Kesteven, Landsberg and URS Australia 2004).

- **Vegetation density** has been shown to be an important determinant of land values in the market place because the value of cleared land tends to be greater than uncleared land (Sinden 2003). A second order term is sometimes included as a potential covariate (vegetation density squared) because there is evidence that land values may increase with vegetation density at lower vegetation levels (Sinden 2003).

- Land values can also be significantly affected by potential erosivity (Palmquist and Danielson 1989). The use of **land capability** data captures erosivity as well as acting as a proxy for other soil productivity characteristics.

- Access to services and businesses that support agricultural industries as well as other lifestyle choices can often be an important determinant of land value. Sometimes referred to as 'distance to town', access is best captured by a proxy measure of **transport cost** from each farm to the closest town of population (say) 5000 or greater (at 2001 ABS Census) along the road network. Transport costs were calculated on a 250 meter grid in a Lambert projection with the cost for each cell consisting of two components:
  - the distance to the closest point on the road network at 16 units per kilometer; and
  - the least cost path along the road network to the closest town of the given minimum size using 8 units per kilometer on sealed roads and 12 units per kilometer on unsealed roads.

- A metric of **frontage to streams** and rivers was also included as a potential covariate to capture productivity benefits associated with stock and domestic water. Measuring effective stream frontages is problematic where streams are heavily folded or in the presence of braided streams – that is, streams that diverge into two or more branches. To overcome these complexities, buffers can be drawn around streams and the ratio of these buffer areas to farm operating area are adopted as the appropriate stream frontage metric. Minor and major streams are assigned buffers of 25 meters and 100 meters respectively.

- The units used to measure farm areas are transformed from hectares to hectares of arable land based on livestock carrying capacities (sheep equivalent per hectare) and cropping land is defined as possessing a carrying capacity of 12 sheep equivalent per hectare. **Land use intensity** is measured as the hectares of arable land relative to the operating area of the farm.

- **Development potential** as measured by the proportion of farm operating area identified for potential grazing or crop development can be included as a potential covariate where it is important to test if landholders' perception of development potential is being capitalized into self-assessed land values.

Amenity values that can be capitalized into land values, such as scenic beauty, are assumed to be zero in broadacre agricultural areas. This assumption seems reasonable given that inland areas largely dependent on commercial agricultural operations and are less likely to be influenced by rural residential demand.

Estimation of the parameters in the land value model equation (1) was by ordinary least squares regression. The model estimates, standard errors and diagnostics summary are reported in table 1.

The regression parameters for the hedonic land value model were tested for robustness using Huber's *M*-estimation (Huber 1981), which made little difference to the results. The model explained around 80 per cent of the variation in the data between farms. All

covariates included in the model had the sign expected and were highly significant.

Table 1: Regression results for log of land value per hectare (All coefficients are statistically significant at the 1 per cent level, $n = 232$, $R^2 = 80\%$)

Variable	Coefficient	Standard error
log(climate)	1.160	0.176
log(travel costs)	-0.445	0.048
log(land use intensity)	0.416	0.041
log(land quality)	-0.242	0.058
stream frontage	0.045	0.009
vegetation density	-0.023	0.004
intercept	3.159	1.190

Farm development potential was not significantly different from zero and accordingly it was concluded that the survey data does not support the hypothesis that farmers were capitalizing development potential into their self assessed land values. Similarly, the data did not support the inclusion of a second order vegetation density term.

6. Concluding Remarks

The linking of scientific and economic data potentially has many benefits for natural resource management and government policy analysis. This paper has presented one very effective and defensible method for bringing these data together in farm surveys.

Analysis presented in this paper indicates that statistical models with very good predictive power can be constructed when scientific spatial data is used together with economic variables for predicting farm values. Similar results have been found for other variables not presented here (Davidson, Elliston, Kokic and Lawson 2005).

In addition, the potential to reduce respondent burden by linking data from other spatial data sources and by a respondent-friendly participatory collection method is a very important benefit of collecting geo-referenced data in farm surveys.

Acknowledgements

The authors would like to acknowledge advice and assistance from Steve Beare, Peter Gooday, Lindsay Hogan, Caroline Levantis, Peter Martin, Colin Mues, Mark Neilsen, Phantipa Puangsumalee, Heidi Rodgers and Vernon Topp of ABARE. The contributions made by Judy Parker and Richard Paton in cleaning the spatial data are much appreciated. The advice and assistance of Greg Rummery of the Walgett Sustainable Agricultural Group Pty Ltd in designing and trialing the survey is particularly appreciated.

This report draws heavily on data collected in ABARE's surveys of broadacre industries. The success of these surveys depends on the voluntary cooperation of farmers, their accountants and marketing organizations in providing data. The dedication of ABARE's survey staff in collecting these data is also gratefully acknowledged. Without this assistance, the analysis presented in this report would not have been possible.

References

ABARE (2004), "Australian Farm Surveys Report," Australian Bureau of Agricultural and Resource Economics, Canberra.

ABS (2006), "Australia and New Zealand Standard Industry Classification," Australian Bureau of Statistics. cat. no. 1292.0.

Bardsley, P., and Chambers, R. L. (1984), "Multipurpose Estimation from Unbalanced Samples," *Journal of the Royal Statistical Society*, *Series C*, 33, 290–299.

Davidson, A., Elliston, L., Kokic, P., and Lawson, K. (2005), "Native vegetation: cost of preservation in Australia," *Australian Commodities*, 12, 543–548.

Huber, P. J. (1981), *Robust Statistics*, John Wiley, New York.

Kesteven, J., Landsberg, J., and URS Australia (2004), "National Carbon Accounting System: Developing a National Forest Productivity Model," Commonwealth of Australia, Australian Greenhouse Office, National Carbon Accounting System technical report; No. 23. http://www.greenhouse.gov.au/ncas

Palmquist (1989), "Land as a differentiated factor of production: a hedonic model and its implications for welfare measurement," Land Economics, 65, 23–28.

Palmquist, R., and Danielson, L. (1989), "A hedonic study of the effects of erosion control and drainage on farmland values," *American Journal of Agricultural Economics* 1, 55–62.

Potgieter, A. B., Apan, A., Dunn, P., and Hammer, G. (2007), "Estimating crop area using seasonal time series of Enhanced Vegetation Index from MODIS

satellite imagery," *Australian Journal of Agricultural Research*, 58, 316–325.

Potgieter, A., Hammer, G., and Butler, D. (2002), "Spatial and temporal patterns in Australian wheat yield and their relationship with ENSO," *Australian Journal of Agricultural Research*, 53, 77–89.

Sinden, J. (2003), "A review of the impacts of vegetation protection on farm land values in northern New South Wales: how much do farmers have to pay?," Submission to the public inquiry "Impacts of Native Vegetation and Biodiversity Regulations" conducted by the Productivity Commission. School of Economics, University of New England, Armidale, NSW 2351.