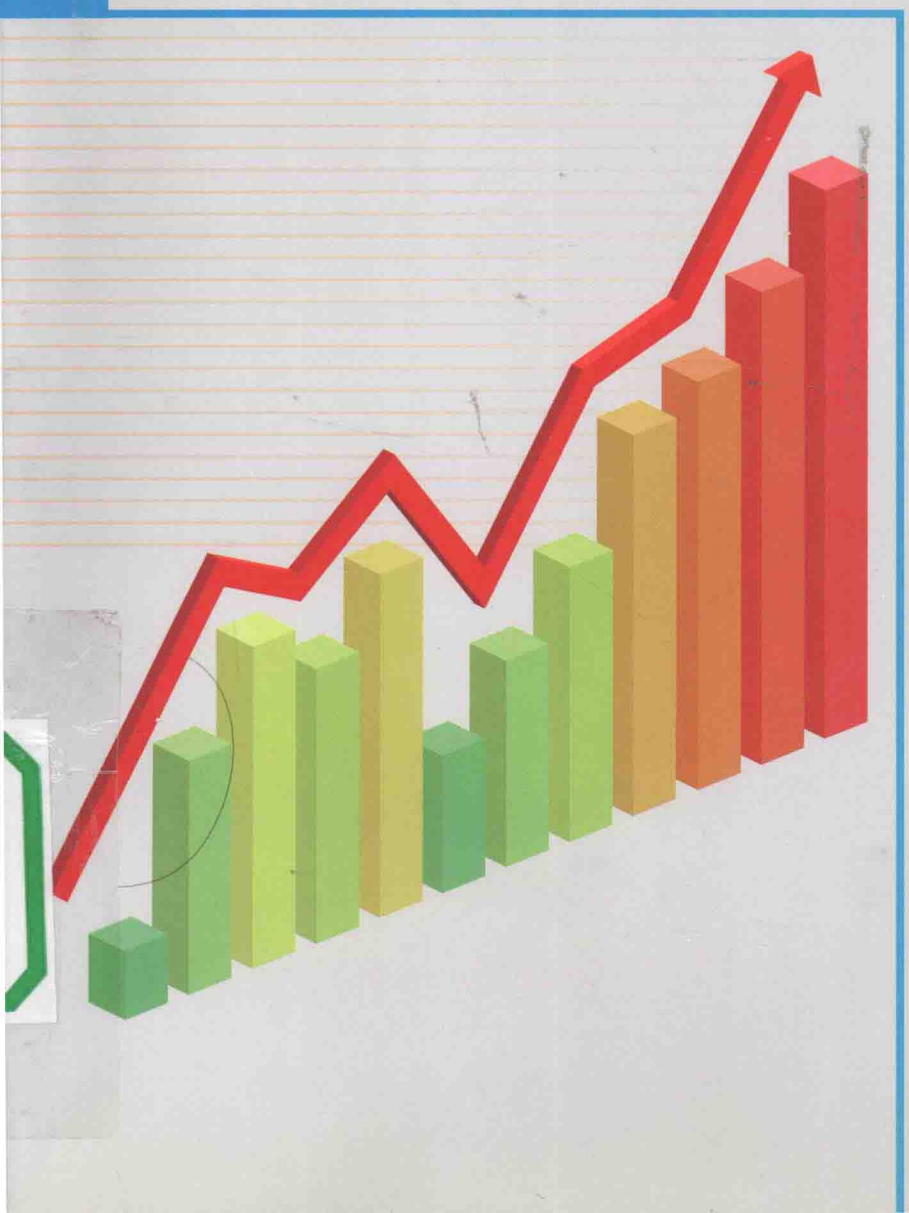
The Application of Remote Sensing and GIS Technology in Crop Production

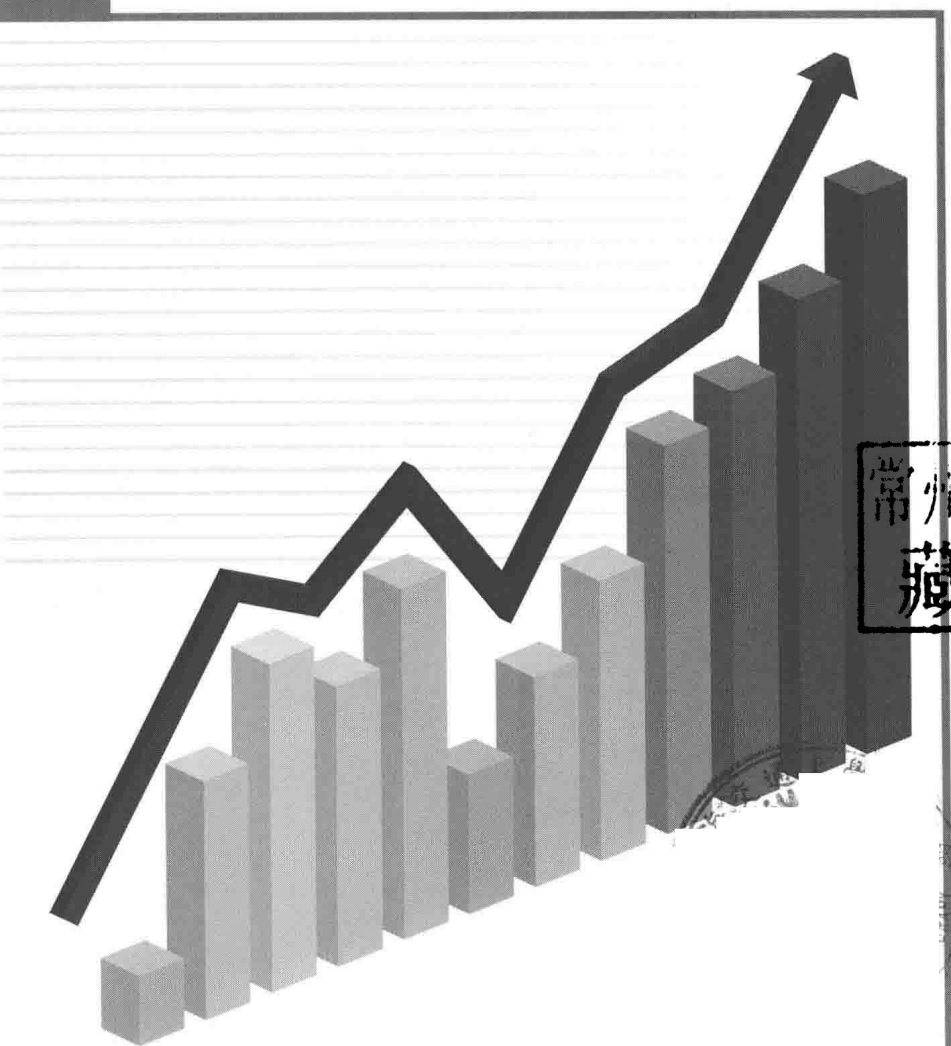
Editors in Chief HE Yingbin, GAO Mingjie, ZHOU Zhenya

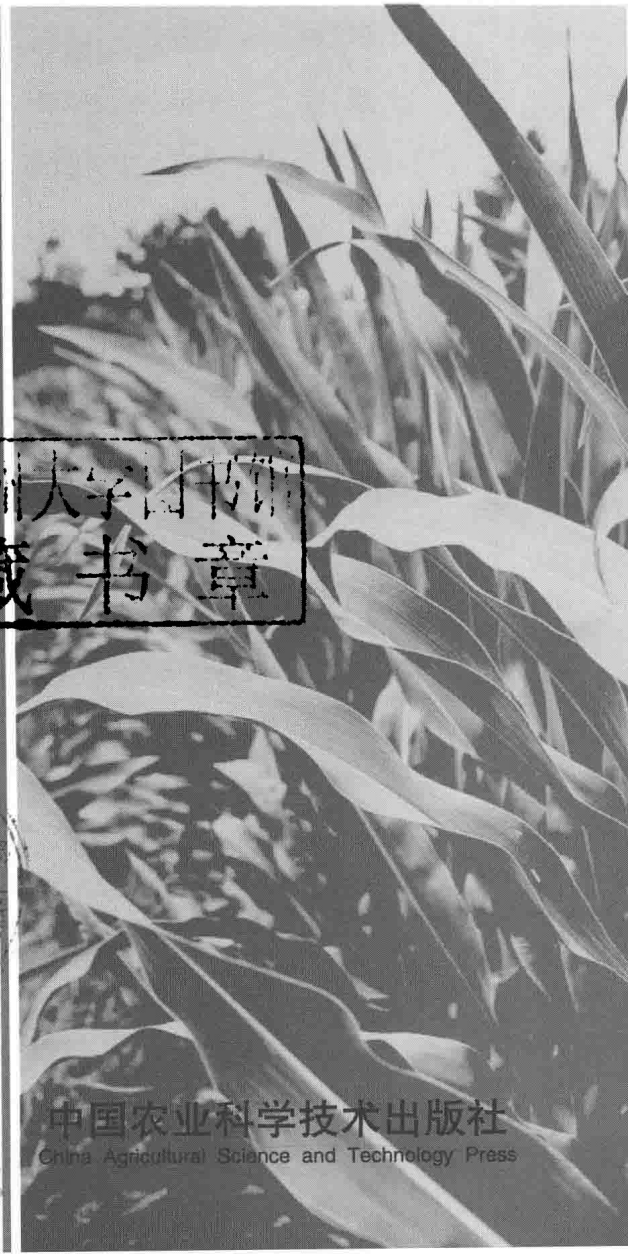


中国农业科学技术出版社
China Agricultural Science and Technology Press

The Application of Remote Sensing and GIS Technology in Crop Production

Editors in Chief HE Yingbin, GAO Mingjie, ZHOU Zhenya





中国农业科学技术出版社
China Agricultural Science and Technology Press

图书在版编目（CIP）数据

遥感与 GIS 技术在作物生产中的应用 = The Application of Remote Sensing and GIS Technology in Crop Production：英文 / 何英彬，高明杰，周振亚主编. —北京：中国农业科学技术出版社，2014.1
ISBN 978-7-5116-1540-4

Ⅰ. ①遥… Ⅱ. ①何… ②高… ③周… Ⅲ. ①地理信息系统—应用—作物—栽培技术—研究—英文 Ⅳ. ①S31

中国版本图书馆 CIP 数据核字(2014)第 018607 号

责任编辑　闫庆健　李冠桥
责任校对　贾晓红
出 版 者　中国农业科学技术出版社
北京市中关村南大街 12 号　邮编：100081
电　话　（010）82106632（编辑室）（010）82109704（发行部）
（010）82109709（读者服务部）
传　真　（010）82106625
网　址　http：//www.castp.cn
经 销 者　各地新华书店
印 刷 者　北京昌联印刷有限公司
开　本　880mm×1230mm　1/16
印　张　8
字　数　300 千字
版　次　2014 年 1 月第 1 版　2014 年 1 月第 1 次印刷
定　价　80.00 元

EDITORIAL COMMITTEE

Prof. ZHOU Qingbo
Deputy Director General, Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences

Prof. LUO Qiyou
Director of Research Branch of Agricultural Allocation and Regional Development, Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences

Prof. CHEN Zhongxin
Director of Research Branch of Agricultural Remote Sensing and Digital Agriculture, Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences

Prof. CHEN Youqi
Research Branch of Agricultural Remote Sensing and Digital Agriculture, Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences

Prof. YANG Peng
Director, Office of Research Administration, Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences

Dr. ZHANG Jizong
Deputy Director, Office of Research Administration, Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences

Mrs. ZHENG Jiang
Office of Research Administration, Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences

MEMBERS (Alphabetized):

AN Xingkui	BAI Jing	BAI Linyan	CAI Weimin
CAO Chenliang	CHEN Youqi	FENG Jianzhong	GAO Mingjie
GUO Bin	HE Yingbin	JIAO Weihua	KONG Qingbo
LI Dandan	LI Guanqiao	LI Runlin	LI Xiangyu
LI Zhibin	LIU Weidong	SHI Shuqin	SI Haiqing
SONG Wenen	WAN Li	WANG Deying	WANG Hao
WANG Huifang	WANG Lei	WEI Dianzhong	XU Xinguo
YAN Hongyan	YAN Qingjian	YANG Fugang	YANG Yimeng
YU Shikai	ZHANG Chen	ZHANG Dongdong	ZHANG Mingwei
ZHNAG Wenbo	ZHOU Zhenya		

Preface

Ensuring food security of ever-increasing population and preserving natural resources are still major challenges for all the APEC economies. New technology, such as remote sensing and GIS technology is shaping economic activity, especially agricultural production, meanwhile, influencing food security. At present, remote sensing and GIS technology has played a key role in assessing and predicting crop acreage and production, monitoring crops growth situation and agricultural disaster. Therefore, we are very much in need of a platform to communicate and share successful experiences, expertise and lessons on application of remote sensing and GIS in crops productivity.

Under this background, and in response to the outputs of a series of ATCWG annual Meeting and 2012 APEC funded workshop, the "APEC Training Course on the Application of Remote Sensing and GIS Technology in Crop Production" was held in August 27th-30th, 2013 in Beijing. The training course aimed at increasing capacity building for APEC economies, especially for developing economies and realizing technical extension in the application of remote sensing and GIS in agriculture among APEC economies. Moreover, the training course was expected to promote advanced technologies adaption among the APEC economies. Therefore, as an important output of the training course, this proceeding will be particularly valuable.

Participants from economies and international organizations presented their research achievements related to the application of remote sensing and GIS technology in crop production during the four-day event. All the presentation were closely linked with crop classification and mapping, crop yield simulation, acreage assessment, crop forecasting, and agricultural diseases monitoring by remote sensing and GIS.

Through the training course, delegates did share applicable and universal method and model in the related field. Moreover, the training course greatly facilitated technological cooperation and improved public awareness in the APEC region.

This proceeding includes 17 papers. And its publication was funded by the China National Natural Science Foundation of China (41001049) and the project "The Training Course of Remote Sensing and GIS Technology in Crop Production" funded by the Chinese Ministry of Finance. The publication of this proceeding is also encouraged by APEC Secretariat, and APEC Budget Management Committee (BMC). I truly express my sincere gratitude to staff working for this training course and all the contributors who helped publish this proceedings. I hope that various people in the field would review this proceedings and that it will contribute to the development of remote sensing and GIS technology.

HE Yingbin

Dec., 2013

CATALOGUE

Accurate Regional to Field Scale Yield Forecasting of Australian Sugar Cane and Peanut Crops Using Remote Sensing and GIS

Andrew Robson[1], Graeme Wright[2]

1. Department of Agriculture, Forestry and Fisheries, Queensland
2. Peanut Company of Australia, Kingaroy, Queensland, Australia, 4610
Ajrob720@yahoo.com.au

Abstract – **The following paper demonstrates the accuracy of satellite imagery for the prediction of average regional yield for both sugar cane and peanut. The development of non- cultivar and non- class specific algorithms that are relatively insensitive to seasonal and location variability offer a useful tool for validating current forecasting methods.The further development of analysis protocols for the rapid derivation and distribution of surrogate yield maps, on mass, offer many applications to their respective industries. These include improved monitoring capabilities to quantify lost productivity resulting from within season incidences of pest and disease outbreaks, as well as unseasonable weather events.**

Keywords– Australian;sugarcane;peanut;yield forecasting

1 INTRODUCTION

Accurate predictions of regional peanut and sugarcane yield are of vital importance for formulating harvesting, milling/ shelling and forward selling decisions. However, if the predictions are made too early in the growing season accuracies can be severely influenced by a wide range of biotic and abiotic constraints. Research conducted under Australian growing conditions has identified that remotely sensed imagery acquired within a growing season can accurately predict the productivity of both peanut and sugarcane. At the regional level, this information has been used to effectively validate pre-season climate modelling based forecasts. At the field scale, the derivation of surrogate yield maps have assisted with improved management strategies, the implementation of precision agricultural technologies and segregated harvesting based of maturity, quality and incidence of diseases including aflatoxin.

Currently, annual production estimates of Australian sugar cane are made early in the cropping season via climate based modelling (Everingham *et al.*2007). Additional yield estimates are made 4 months into the season through the visual assessment of selected crops. These methods although generally accurate, do not account for mid to late season climatic anomalies such as large rain events and associated flooding, or outbreaks of pest and disease.

For peanut, estimates of annual production are derived from the amount of seed distributed at the start of the season. Although, this method provides reasonable accuracies in some years, it does not take into account yield and quality fluctuations arising from variations in seasonal conditions and planting rates.

Geographic information systems (GIS) and remote sensing (RS) technologies offer an additional tool for validating these current methods of prediction.

2 METHODOLOGY

2.1 Study Districts

Sugarcane: Research was conducted across three climatically distinct Queensland cane growing regions, the Herbert, the Burdekin and Bundaberg / Isis during the 2010, 2011 and 2012 growing seasons. Comprehensive Geographical Information System (GIS) vector layers of all crops grown within each growing region and for each season were obtained from the relevant industry groups.

Peanut: Research was conducted over dryland and irrigated cropping regions of Queensland including the South Burnett, Atherton and Bundaberg as well as Katherine in the Northern Territory. This research has been continually developed since 2004. Unlike the cane industry comprehensive GIS boundary layers defining Peanut crops and associated attribute data have not been established. Therefore crops were defined through the supervised classification of SPOT5 and QuickBird imagery, on-ground validation and then digitization to develop crop boundaries.

2.2 Development of the yield prediction algorithm: Sugarcane

The French owned SPOT5 satellite was identified to be the most appropriate for the Australian sugar industry due to its spatial and spectral resolutions, cost, revisit time and minimum scene area of 3600 km^2. To convert spectral values into tonnes of cane per hectare (TCH), a non- cultivar, non- crop class specific algorithm was derived from the correlation between average cane yield and corresponding average crop greenness normalised difference vegetation index (GNDVI) value. GNDVI was selected as it expressed less saturation with large canopy cane crops than the commonly used NDVI.

$$GNDVI = (P_{NIR} - P_{GREEN}) / (P_{NIR} + P_{GREEN}) \quad (1)$$

Where P_{GREEN} and P_{NIR} are the Top of Atmosphere corrected reflectance values measured in the green and near infrared SPOT 5 spectral bands

GNDVI yield algorithm (Bundaberg)

$$y = 3.1528 * e^{(5.6973 * x)} \quad (2)$$

Where y = predicted average yield (TCH) and x = average GNDVI value extracted from TOA SPOT5 image ($r = 0.77$, $n = 151$ crops).

An additional algorithm was derived for the Burdekin region due to its vastly different growing environment.

GNDVI yield algorithm (Burdekin)

$$y = 12.691 * e^{(3.8928 * x)} \quad (3)$$

Where y = predicted average yield (TCH) and x = average GNDVI value extracted from TOA SPOT5 image (r=0.42, n= 4573 crops).

For a complete methodology of the sampling procedure used refer to Robson, A., Abbott, C., Lamb, D., and Bramley, R. (2012).

2.3 Development of the yield prediction algorithm: Peanut

Due to the smaller cropping area of Peanut within Australia, the yield prediction algorithm was derived from imagery acquired from the American owned QuickBird satellite. The algorithm was developed from the correlation between NDVI and peanut yield in tonnes per hectare (t/ha), measured at 295 locations. The crops were sampled between 2004 and 2009 and included 10 cultivars from both dryland and fully irrigated systems within Queensland and the Northern Territory. For a complete methodology of the sampling procedure used refer to Robson, A.J. (2007).

$$NDVI = (P_{NIR} - P_{RED}) / (P_{NIR} + P_{RED}) \quad (4)$$

Where P_{RED} and P_{NIR} are the red and near infrared QuickBird spectral bands.

NDVI Peanut yield algorithm

$$y = 0.2104 * e^{(4.455 * x)} \quad (5)$$

Where y = predicted average Peanut yield (t/ha) and x = average NDVI value extracted from Quickbird imagery (r=0.74, n= 351 crops).

2.4 Regional Predictions of Sugarcane Yield and generation of yield maps

A SPOT5 satellite scene was captured over each of three target regions during 2010, 2011 and 2012, with the exception of a 2010 Herbert image due to continual cloud cover. Image acquisition was between March and May, a period between vegetative growth and senescence where sugarcane exhibits stabilization in development and NDVI (Bégué *et al.* 2010; Almeida *et al.* 2006).

All SPOT5 images were corrected for top of atmosphere reflectance (TOA) (SPOT Image, 2008) and orthorectified. Within ArcGIS, a 20 m metre internal buffer was applied to the boundary of each individual field to ensure the extracted spectral information did not include non cane-specific pixels. The extraction of spectral information was undertaken with the software StarSpanGUI (http://projects.atlas.ca.gov/projects/starspan/). This data along with corresponding attribute mill data for each field was exported into Microsoft Excel for additional analysis. The average GNDVI for all crops was calculated and then substituted into the appropriate yield algorithm producing a predicted average regional yield. This prediction was then validated against the actual reported yield provided by industry, following harvest.

For the derivation of yield maps, the TOA georectified SPOT5 images were converted into a GNDVI layer using ENVI (http://www.exelisvis.com/).The accompanying GIS boundary files were transformed into separate regions of interest (ROI's) and used to extract the spectral information of all sugarcane crops. Using the ENVI band math function, all GNDVI pixel values were converted to TCH, using the appropriate algorithm. The crops where then classified into 8 yield classes using a density slice, and saved as a Geotiff layer or within a GoogleEarth KMZ file for distribution to end users.

2.5 Regional Predictions of Yield and generation of yield maps. Peanut

A similar process as that defined for sugarcane was used for Peanut, with the exception that a 5 metre internal crop buffer was used instead of the 20 metre for sugarcane. This was due to the 2.4 metre spatial resolution of non- pan-sharpened QuickBird imagery compared to the 10 metre SPOT5 imagery.

The mean NDVI value for each crop was calculated, using the ENVI image "statistics" option, while the respective areas (hectares) were measured using ArcMap 10 (UTM, WGS-84). The average NDVI value was applied to the exponential equation, producing an average yield for the each crop (t/ha). The average yield was then multiplied by the area of the crop (ha) to produce a prediction for total yield. To validate the accuracy of the prediction, the total yield for each crop was compared to the actual payable harvested weight of farmer stock delivered to peanut processors. This involved additional QuickBird satellite images being captured over a number of intensive dry land and irrigated peanut cropping areas from 2009 to 2012.

3 RESULTS

3.1 Prediction accuracies of the sugarcane algorithms at the regional level

To assess the predictive accuracy of the algorithm at a regional level, estimates of average yield were compared to those provided by the Mills following harvest. As seen in Table 1, the predicted average yields were highly comparable to the actual values, with the exception of 2010 Bundaberg and Isis.

The further derivation of yield maps for all crops imaged (Figure 1) allowed for regional trends in productivity to be easily identified. The prediction accuracies at the block level did display varying degrees of accuracy, but as seen in Figure 2 has the potential to be highly accurate. Considering the non cultivar and non class specific nature of the Bundaberg and Burdekin algorithm, this result was highly encouraging.

Table 1 Predicted versus actual average yield (TCH) from 4 growing regions across 3 growing season

Harvest year	Growing Region	Number of crops	Pred.ave.Yield (TCH)	Act.ave.Yield (TCH)
2010	Bundaberg	3544	79.7	81.8
2011	Bundaberg	3824	80.1	73.3
2012	Bundaberg	3217	88.0	88.9
2010	Isis	2772	84.0	84.0
2011	Isis	4205	98.4	83.3
2012	Isis	4000	92.5	96.0
2011	Herbert	8596	51.4	55.0
2012	Herbert	15463	75.0	72.0
2011	Burdekin	4999	118.8	120.0
2012	Burdekin	6921	110.0	105.0

3.2 Prediction accuracies of Peanut algorithm at the regional and crop level

The validation of the peanut yield prediction algorithm on a number of irrigated and dry land crops grown across Australia produced high prediction accuracies. The example provided in Figure 3, indicates the predicted yield of 849 tonnes from a 317 ha dry land crop (cv. Walter), was highly comparable to the actual payable yield of 874 tonnes.

Prediction accuracies for irrigated crops were less than those achieved for the dryland crops, around 85% - 90% of actual yield. A result attributed to a NDVI saturation point occurring around 0.83, or yield prediction sensitivity above 8 t/ha. This was further exacerbated by the effect of regional growing environment especially that in the Atherton tableland. An additional algorithm was therefore derived for this region(Figure 3).

Using the Atherton specific algorithm, the average Peanut yield predicted for 75 crops grown during the 2012 season was 5.1 t/ha, the actual yield was reported at 4.9 t/ha. The spatial variability of yield was again derived at the field and regional level by converting the NDVI values into yield (Figure 4). As with the derived yield maps for sugar cane, these surrogate maps were provided to industry via a range of media.

4 CONCLUSIONS

The results presented in this paper demonstrate many beneficial applications to both the Peanut and Sugar cane industries. At the regional level, the accurate yield forecasts offer an important tool for validating and enhancing current forecasting methods; whilst the generation of yield maps on mass offers an accurate insight into sub regional trends in production. By having access to multiple years of these maps, rapid changes in sub-regional trends, including climatic effects of flooding or drought stress, as well as outbreaks of pest and disease can be identified.

For peanut access to this information has also assisted with crop auditing for claiming royalties associated with plant breeding rights. At the field level, the surrogate yield maps have assisted with harvest segregation in response to aflatoxin risk. Future research will see the development of additional yield prediction algorithms from SPOT5 imagery. This will enable yield maps for mixed peanut and sugar cane farming systems to be generated from the one image capture.

For the sugar cane industry, 2013 will see coverage extended to the majority of the Australian growing regions encompassing 9 SPOT 5 images and over 40000 fields. Additional biometric analysis of data sets will also be undertaken to further improve the prediction accuracies at the block level. This will involve the development of cultivar specific algorithms.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the Sugar Research and Development Corporation (SRDC) for providing funding for this research as well as those growers and industry partners whom have collaborated. The authors would also like to acknowledge the Peanut Company of Australia and the University of Queensland for funding assistance for the Peanut research.

REFERENCES

[1] Almeida, T., Filho, C., Rossetto, R.2006. ASTER and Landsat ETM+ images applied to sugarcane yield forecast. International Journal of Remote Sensing 27, 4057-4069.

[2] Bégué, A., Lebourgeois, V., Bappel, E., Todoroff, P., Pellegrino,

A., Baillarin, F., Siegmund, B. 2010. Spatio-temporal variability of sugarcane fields and recommendations for yield forecasting using NDVI. International Journal of Remote Sensing, 31(20).5391- 5407.

[3] Everingham, Y.L., Inmam-Barber N.G., Thorburn P.J., McNeill J.T. 2007. A Bayesiasn modelling approach for long lead sugarcane yield forecasts for the Australian sugar industry. Australian Journal of Agricultural research, 58 (2), 87-94.

[4] Andrew, R., Chris, A., David, L., Rob, B. 2012. Developing sugar cane yield algorithms from satellite imagery. Proceedings of the Australian Society of Sugar Cane Technologists. 34th Conference, Cairns, Qld. 1- 4 May.

[5] A.J.,Robson. 2007. Remote sensing applications for the determination of yield, maturity and aflatoxin contamination in Peanut. Ph.D. Thesis, the University of Queensland.

[6] SPOT Image .2008.SPOT Image Homepage April 28th.Website link http://wwwspotimagecom.

An Empirical Analysis of Land Use Change and Soil Carbon Sequestration in China

Li Man [1,*], Wu Junjie [1], Deng Xiangzheng [2]

[1] Department of Agricultural and Resource Economics, Oregon State University, Corvallis, OR 97331, USA

[2] Center for Chinese Agricultural Policy, Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China

manlichn@gmail.com

Abstract-**This paper presents an empirical analysis of land use change from 1985 to 2005 and its impact on soil carbon storage in China. It compiles a national-level geographic information system (GIS) database that includes data on land use, soil, and climate. It also develops a statistical method for evaluating the impact of land use change on soil organic carbon storage. The method greatly reduces data requirements for policy analysis at the national level. Results indicate that land use change from 1985 to 2005 caused a total increase in topsoil carbon stock by about 7.5 TgC or 0.02 percent. The largest soil carbon increase occurred when sparser grasslands were converted to denser grasslands, while the largest soil carbon loss occurred when denser grasslands were converted to sparser grasslands.**

Keywords-China; land use change; soil carbon sequestration

***JEL classification*: Q15, Q24**

1 INTRODUCTION

Soil is the largest carbon pool in global terrestrial ecosystems. Soil carbon sequestration, defined as using and managing land in ways that enhance the natural adsorption of atmospheric carbon by soil, is considered as a win-win strategy because it not only mitigates greenhouse gas (GHG) emissions but also generates economic benefits by improving soil fertility and agricultural productivity (Antle et al. 2007; Bauer and Black 1994; Feng et al. 2006, 2007; Lipper and Cavatassi 2004; McCarl and Sands 2007).

The total amount of soil organic carbon (SOC) in China was estimated in the range of 69.1 to 92.4 PgC (1 PgC = 10^{15} gram carbon) down to the depth of one meter (Wang et al. 2001; Wang et al. 2004; Wu et al. 2003; Yang et al. 2007). The findings of spatial distribution of SOC density are generally consistent in the literature. The density decreases from the southeast to the northwest while increases from arid to semi-humid zone in northern China and from tropical to cold-temperate zone in eastern China. The carbon density also decreases with soil depth. Approximately 46–54 percent of soil carbon is stored in the upper 30 centimeters (Wang et al. 2004; Wu et al. 2003; Yang et al. 2007), almost three times as many as carbon stocks in vegetation (Piao et al. 2004; Piao et al. 2005).

The long-term storage potential of soil carbon is limited by site characteristics such as soil type and climate (IPCC 2006). The extent to which that storage potential is realized, however, depends largely on how land is used and managed (Bruce et al. 1999; Houghton 2003, Paustian et al. 1997; among many). The literature on soil carbon sequestration in China is extensive; many facets of agricultural land management have been examined for their potential to increase soil carbon (Huang and Sun 2006; Lal 2002; Tang et al. 2006; Zhang et al. 2007; Zhang et al. 2006). Only a few studies, however, have considered changes in soil carbon stocks in multiple ecosystems including forests and grasslands (Ge et al. 2008; Piao et al. 2009; Wang et al. 2003).

China has experienced tremendous land use changes in the last few decades, characterized by three distinguished features: (1) rapid urban expansion into fertile farmland in the North China Plain, the Yangtze River Delta, and the Sichuan Basin, (2) cultivatable land reclamation on woody areas in Northeast China and southeast hilly regions and on grassland in Northwest and Central China, and (3) bi-directional conversions between woody and grass lands in southeast coastal and southwest regions (Liu et al. 2003). Besides, China witnessed about 360 thousand hectares of grass and arable lands annually encroached by the Gobi Desert (The Gobi is a large desert region in Asia. It covers parts of northern and northwestern China, and of southern Mongolia) because of overgrazing and overplowing. This has led to a loss of topsoil of 200 thousand hectares each year (Steffen 2003).

The purposes of this study are twofold. First, it develops a statistical method to estimate the effect of land use change on soil carbon densities. Second, it assesses changes in soil carbon storage caused by land use change from the 1980s to the 2010 decade. To achieve these objectives, we compile a geographic information system (GIS) database that contains three datasets. The first one includes high-resolution GIS land use data for four time periods—the mid-1980s, the mid-1990s, the late 1990s and the middle years of the 2000–2010 decade, denoted as 1985, 1995, 2000 and 2005, respectively. The land use data were originally derived from the U.S. Landsat Thematic Mapper scenes with a 30-meter spatial resolution.

Landsat images were interpreted by the Data Center for Resources and Environmental Sciences, Chinese Academy of Science (CAS) with a hierarchy land cover classification system and were validated by extensive ground-based surveys (Liu et al. 2003). The average interpretative accuracy is more than 97 percent (Liu and Buheaosier 2000; Liu et al. 2003). The second dataset includes cross-sectional SOC estimates and soil classes that were initially derived from the Harmonized World Soil Database (FAO/IIASA/ISRIC/ISS-CAS/JRC 2009). The third one is the Köppen-Geiger climate classification map (2006) that is one of the most widely used climate classification systems.

The long-term SOC storage is determined by the balance of litter inputs from plant production and carbon emissions through a decomposition process (Jobbágy and Jackson 2000; Parton et al. 1993; Schlesinger 1977), and the impact of land use change on SOC depends on the site-specific characteristics such as land quality and weather conditions. Thus, to evaluate the effect of a conservation policy on SOC, it is necessary to determine how the policy affects land use and how changes in land use in turn affect SOC at each location. This is typically accomplished through a process-based simulation model calibrated for a specific site or watershed. It is practically infeasible, however, to simulate environmental impacts at all sites and for all sets of conditions that arise in a national analysis such as the one performed here.

This study makes a methodological contribution by developing a statistical method for evaluating the impact of land use change on soil organic carbon storage at the national level. Instead of relying on simulations, our method takes advantage of cross-sectional variations in land use and soil carbon density and uses the information to identify the relationships between land use change and SOC density. This approach can be applied to a large region and hence overcomes the limitation of a process-based model that typically works for a relative small region.

Previous studies, particularly those based on field measurements and soil inventories, are often conducted at the regional level or focus on a specific soil type. Ge et al. (2008) adopted a bookkeeping method to analyze the effects of land use and land cover change on the terrestrial carbon balance at provincial level for the period 1700–1949. Wang et al. (2003) investigated patterns and changes in SOC storage by region from the 1960s to the 1980s based on China's first and second national soil surveys. Piao et al. (2009) assessed the carbon balance (i.e., annual carbon change) of China terrestrial ecosystems between the 1980s and 1990s by integrating biomass and soil inventories with the remote-sensing NDVI (normalized difference vegetation index) data. Huang and Sun (2006) conducted a meta-analysis using data from 132 publications to evaluate topsoil carbon changes on cropland. None of the previous studies, however, has evaluated the effect of land use and conservation policies on SOC in China at the national level.

The remainder of this paper is organized as follows: Section 2 describes materials and methods. Section 3 discusses the results. The final section discusses the main findings of this study.

2 MODELING THE EFFECT OF LAND USE CHANGE ON SOC DENSITY

This section describes the method for evaluating the effect of land use change on SOC density. Following the IPCC guidelines for SOC inventories, we make the following assumptions in the evaluation: (1) Soils tend towards equilibrium under a given set of climate and soil conditions; (2) The stock of SOC changes in a linear fashion during the transition to a new equilibrium. Assumption 1 has been widely accepted in the literature; assumption 2 simplifies the estimation of SOC changes and provides a good approximation of SOC changes over a multi-year period (IPCC 2006, pp. 2.29).

Under these simplifying assumptions, change in SOC density, ΔC, between time 0 and T can be calculated as:

$$\Delta C = (C^1 - C^0) \times T/\bar{T} \quad (1)$$

Where C^0 is carbon density at the beginning of land use transition period, C^1 is carbon density when soils achieve new equilibrium, and $\bar{T}$ is the number of periods it takes to reach the new equilibrium. We set T equal to $\bar{T}$ for $T > \bar{T}$.

It would be easy to calculate the change in SOC density when C^0 and C^1 were known. But multiple soil inventories are rarely available. In practice, C^1 can be extrapolated using one-period soil inventory and some stock change factor. Specifically, given C^0, C^1 can calculated by

$$C^1 = C^0 \times \exp(\delta_{jks}) \quad (2)$$

Where δ_{jks} is the stock change factor in response to a specific land use conversion from j to k stratified by a combination of soil and climate class s. Taking the natural logarithm on both sides of (2), we have

$$ln\, C^1 = ln\, C^0 + \delta_{jks} \quad (3)$$

Practically, stock change factor is often derived from experimental data, by estimating the factors from each study or observation and then analyzing those values using appropriate statistical technique (e.g., Ogle *et al.* 2004). Since studies on changes in soil carbon in China's multi-ecosystem are generally lacking, we use a treatment effect analysis to estimate δ_{jks} based on a digital SOC density map and remote-sensing land use data.

Assume in a *natural* experiment, land use represents a group of mutually exclusive treatments and a log-transformed SOC density is the subsequent outcome. A cause is viewed as land conversion from one use (hereafter control) to an alternative use (hereafter treatment) that brings about a change in the log-transformed SOC density. Given any land unit n that is exposed to a control j, the causal relationship of land use conversion and SOC density can be expressed as

$$(ln\,C)_{jksn} = \mu + \tau_{jk} + \rho_s + (\tau\rho)_{jks} + \varepsilon_{jksn} \quad (4)$$

Where μ is the general mean, τ_{jk} is the treatment effect of changing land use from j to k, ρ_s is the soil-climate block effect, $(\tau\rho)_{jks}$ is the treatment × block interaction, and ε_{jksn} is a random error with mean 0 and variance σ^2. Given any soil and climate condition s, a conversion from a control use j to a treated use k would cause $\tau_{jk} + (\tau\rho)_{jks}$ unit changes in the logarithm of SOC density when achieving the new equilibrium (that is, $\delta_{jks} = \tau_{jk} + (\tau\rho)_{jks}$). The underlying assumption is that soil temperature, moisture, and texture interact with land use activities, jointly controlling carbon absorption and release mechanisms.

2.1 The Estimation Method

Spatial autocorrelation is one econometric consideration because SOC density in neighboring land units may have unobserved factors that are correlated over space, leading to inefficient estimates and invalid hypothesis testing procedures. There are two ways to control such correlation; one is to specify a non-spherical error variance-covariance matrix and is often called a spatial error model (Anselin 1988), and the other is to eliminate nearest neighbors from the sample, commonly called spatial sampling routine (Besag 1974; Haining 1990). The advantage of the first approach is that when properly specified, it can generate efficient estimates, but the approach could make computation infeasible when data sets are large, because the estimation relies on the maximum likelihood principle (Magnus 1978; Mardia and Marshall, 1984; among many). In addition, the estimation results are subject to a specific and restrictive set of assumptions of non-spherical error variance-covariance matrix. The second approach will not, in general, be fully efficient but its great advantage lies in its simplicity and flexibility. This approach has been demonstrated effective at reducing or eliminating potential spatial autocorrelation in the error terms in many land use studies.

The basic unit of observation in this study is 1 km × 1 km pixel, of which there are more than 9.5 million to cover the entire country. Therefore, we choose the second approach to control spatial autocorrelation; we take a 1-out-of-25 sample by choosing only the pixels at the centers of a 5 km× 5 km grid. It would allow us to take advantage of the large sample while avoiding intensive computation.

Another important concern is the nonrandomized data used in this study. Unlike a randomized experiment in which the outcomes from treatment and control groups are often be directly compared because their units are generally similar, a nonrandomized experiment may produce biased estimates when directly comparing a treatment group to a control group, because the units exposed to a treatment could differ systematically from the units exposed to a control. This is often called sample selection (or self-selection) bias in the literature.

To correct for such potential bias, a number of matching estimators have been proposed in the econometric/statistical literature(See Imbens 2004 for a thorough review). The idea of matching is, in sampling from a large reservoir of potential controls to pair the treated and control units that are similar in the distribution of covariates, the receipt of treatment is likely to be independent of the outcomes with and without treatment (Rosenbaum and Rubin 1983).

In this study, we follow the strand of nearest neighbor matching and take two strategies to make the treatment and control groups more comparable to each other. First, we stratified land units into soil-climate blocks in the conceptual model,see equation (4). Within each block, land use treatments can be compared under a relatively uniform environment. Second, we pair the treated and control units that are geographically closest to each other. The underlying assumption is that nearby land units are more likely to have similar characteristics, including topographic position, soil physical properties, and other unobserved features. Matching is implemented after spatial sampling.

One issue arises in implementing matching is whether or not to match with replacement. Matching with replacement minimizes the Euclidean distance between the treated units and the matched control units, which is beneficial in terms of bias reduction. However, when only few control units close to the treated units, the control units will be repeatedly used as a match. This increases the variance of the estimates. Matching without replacement improves the precision of the estimates but it could raise bias; it would also complicate the algorithm because the results are potentially sensitive to the order in which the treatment units are matched (Rosenbaum 1995).

A closely related question is whether to consider the order of treatment and control, i.e., matching each treated unit to a control or reversely, matching each control unit to a treated. In the former case, the focus is on the subpopulation of treated units and the derived estimator is actually the estimated treatment effect when changing land use from the control in the

past to the treatment in the present. In the latter case, of more interest is the subpopulation of control units and the derived estimator is the estimated treatment effect of conversion from the current control use to the future treatment use.

The selection of matching strategy is an empirical question. It depends on the geographical distance between the treatment and control groups. When most units in the treatment and control groups are geographically close to each other, the foregoing methods will yield similar results. In the application that follows, we consider three matching estimators: the treatment effect for the treated population, the treatment effect for the control population, and the treatment effect for overall population.

Specifically, given any pair of treatment and control groups, we begin with matching each treated unit to a control unit to produce the first estimates (treatment effects for the treated population). In this case, every treated unit is used only once but the control can be repeatedly selected. We then match the pair in an opposite direction and generate the second estimates (treatment effects for the control population). In this sample, the treated units, instead of control, can be chosen more than once. Note that the order of treatment and control matters when producing these two estimates. In particular, for each pair of treatment and control groups, the first estimates are exactly equal to the second estimates produced by interchanging the order of treatment and control.

Finally, we overlap the two samples generated by using the first and second approaches, and use their intersection to produce the third estimator (treatment effects for overall population). By doing so, we pair the treated and control units that are truly nearest neighbors, which could further reduce the potential bias from sample selection. With this approach, both treated and control units are used only once and the treatment effects are symmetric.

2.2 Data

Our study uses a GIS database that covers whole China, including a digital SOC density map, land use data for four years, IPCC soil class map, and Köppen-Geiger Climate Classification map.

2.2.1 SOC Density

The digital SOC density map was generated from Global Soil Organic Carbon Estimates that were developed by Joint Research Center, European Commission (Hiederer and Köchy 2012), designed to support the calculation of potential emissions of CO_2 from the soil at a finer resolution (1 km) under the IPCC land use and climate change scenarios. The stocks of SOC were computed separately for the topsoil (0–30 cm) and subsoil (30–100 cm) layers using SOC content, gravel content, soil depth, and bulk density data in the Harmonized World Soil Database (HWSD, version 1.1). In particular, the spatial layers on parameters of soil in China were initially derived from the Soil Map of China based on data of the office for the Second National Soil Survey of China (1995) and were distributed by the Institute of Soil Science in Nanjing (Shi et al. 2004). This Survey was a special nationwide research and documentation project organized by the State Council and carried out by a consortium of universities, research institutes, and soils extension centers. It includes soil inventory data from 2,473 typical soil profiles collected during 1979–1985.

We use the layer of top 30 cm for the presented analysis as soil carbon stored in the topsoil is most likely to be affected by land use changes. Figure 1 maps the spatial distribution of China's topsoil carbon density.

2.2.2 Land Use

CAS generated the contiguous land use data based on the U.S. Landsat Thematic Mapper/Enhanced Thematic Mapper (TM/ETM) images (Liu and Buheaosier 2000; Liu et al. 2003; Liu et al. 2010). The data are available for four time periods—the mid-1980s, the mid-1990s, the late 1990s, and the middle years of the 2000–2010 decade—denoted as 1985, 1995, 2000, and 2005, respectively. There are more than 500 TM scenes for each period. CAS made visual interpretations and digitization of TM/ETM images to derive thematic maps of land use percentages with a spatial resolution of 1 km and sorted the data with a hierarchical classification system of 25 land use classes, which were further grouped into nine aggregated classes: paddy, dryland, woodland, dense grassland, medium grassland, sparse grassland, water area, urban land, and unused land (Figure 2). In particular, woodland includes natural and planted forests and land used for tea-gardens, orchards and nurseries; water area is classified as land covered by natural water bodies and land with facilities for irrigation and water conservation; urban land includes land used for urban and rural settlements, industry, and transportation. A detailed explanation of the nine aggregated land use classes is available in the appendix.

Table 1 depicts land use conversions, by assigning each land unit to the use with the highest proportion, among these classes for 1985–2005, where entries in a cell indicate the number of Mha that were in the row land use in 1985 and column land use in 2005. The entries along the diagonal are areas where land use has not changed. All land uses except grasslands and water area increased. Land use changes occurred mainly from farmlands (paddy and dryland) to urban