Naik, G., Basavaraj, K.P., Hegde, V.R., Paidi, V. and Subramanian, A. (2013) Using geospatial technology to strengthen data systems in developing countries: the case of agricultural statistics in India. *Applied Geography*, 43, pp. 99-112. (doi:10.1016/j.apgeog.2013.06.002)

This is the author’s final accepted version.

There may be differences between this version and the published version. You are advised to consult the publisher’s version if you wish to cite from it.

http://eprints.gla.ac.uk/84821/

Deposited on: 29 August 2017

Enlighten – Research publications by members of the University of Glasgow
http://eprints.gla.ac.uk33640

Using geospatial technology to strengthen data systems in developing countries: The case of agricultural statistics in India

Gopal Naik[a], K.P.Basavaraj[a], V.R. Hegde[b], Vijay Paidi[a] & Arjunan Subramanian[c]

[a] Indian Institute of Mangement, Banglore, India
[b] Zoomin Softech, Bangalore, India
[c]University of Glasgow, Glasgow, UK

Contact person:

Dr. Arjunan Subramanian
Economics
Adam Smith Business School
Gilbert Scott Building (Room 407b)
University Avenue
University of Glasgow
Glasgow, G12 8QQ
Scotland
UK
T: +44 (0)141 330 7304
F: +44 (0)141 330 4940
Email: arjunan.subramanian@glasgow.ac.uk

Acknowledgement: Previous versions of this paper were presented as invited paper in the National Workshop on Land Records Management in India: A Plea for Reform, Hyderabad (India), January 23-24, 2012 and the Scottish Institute for Research in Economics Young Researchers Forum (SIRE) in the University of Strathclyde Business School, Glasgow (UK), October 31, 2012. The authors thank all the participants for their valuable comments. Thanks are also due to Tony Gloyne for extensive comments on a previous version of this paper. One of the authors acknowledge funding from the Economic and Social Research Council for the project "Information, market creation and agricultural growth" (ES/J009334/1) under the ESRC DFID Growth Programme and also the support of the Royal Society of Edinburgh under the RSE International Exchange Programme.

Using geospatial technology to strengthen data systems in developing countries: The case of agricultural statistics in India

ABSTRACT. Despite significant progress in the development of quantitative geography techniques and methods and a general recognition of the need to improve the quality of geographic data, few studies have exploited the potential of geospatial tools to augment the quality of available data methods in developing countries. This paper uses data from an extensive deployment of geospatial technology in India to compare crop areas estimated using geospatial technology to crop areas estimated by conventional methods and assess the differences between the methods. The results presented here show that crop area estimates based on geospatial technology generally exceed the estimates obtained using conventional methods. This suggests that conventional methods are unable to respond quickly to changes in cropping patterns and therefore do not accurately record the area under high-value cash crops. This finding has wider implications for commercializing agriculture and the delivery of farm credit and insurance services in developing countries. Significant data errors found in the conventional methods could affect critical policy interventions such as planning for food security. Some research and policy implications are discussed.

Keywords: Agriculture, poverty, geospatial, global positioning system, geographic information system

JEL: C81; R12; R14; Q15

Introduction

The recent World Development Report on "Agriculture for Development" recognized that agriculture is central to achieving the Millennium Development Goals of poverty reduction and environmental sustainability (World Bank, 2008). However, the quality of available agricultural data and the methods by which such data are collected are notoriously weak in several developing countries. Recent developments in quantitative geography offer robust geospatial tools that provide access to new data and methods for strengthening data systems (Bell & Dalton, 2007; Murray, 2010). Surprisingly, despite the availability of these new tools and methods in these countries, their application has been limited (Miller, 2010). This paper addresses this gap by examining the current data systems and demonstrating the significant potential role for geospatial tools in improving the quality of agricultural data and the methods by which it is obtained and thereby permitting better policy in developing countries.

Although there is general recognition of a longstanding need for strengthening agricultural data availability in developing countries (United Nations, 1979; World Bank, 2011; African Development Bank Group, 2011), surprisingly little research exists on the reliability of agricultural data and the methods by which such data are collected (Beegle, Carletto, & Himelein, 2012; Deininger, Carletto, Savastano & Muwonge, 2012). One exception is Muller, Muller, Schierhorn & Gerold (2011), who used spatiotemporal data to study the dynamics of deforestation attributable to mechanized agriculture. Recognizing the lack of geospatial data on land use and land cover in developing countries, Dewan & Yamaguchi (2009) use data from Bangladesh to analyze the spatial and temporal characteristics of urban land expansion. In contrast, developed countries use more advanced geospatial tools that combine global positioning systems with video for field data collection (Mills, Curtis, Kennedy, Kennedy & Edwards, 2010).

Some recent studies have examined the reliability of household consumption data in India (Sen, 2000; Kulshrestha & Kar, 2005; Deaton & Kozel, 2005) and Tanzania (Caeyers, Chalmers & Weerdt, 2012), but research on the quality of data on the production side of agriculture remains limited. We are aware of only two recent contributions examining the reliability of traditional recall-based survey methods in the generation of agriculture production statistics. The evidence from these studies is mixed; while Beegle, Carletto, & Himelein (2012), using data from three African countries, found little evidence of a large recall bias in agricultural data, Deininger, Carletto, Savastano & Muwonge (2012), in contrast, noted significant differences between data generated from recall-based surveys and data from production diaries administered concurrently in Uganda. However, it is not clear yet which of these two methods yields the more accurate results.

An advantage of using crop area statistics to examine data reliability is that crop area is both measureable and independently verifiable using existing technology. We use data from India, which has one of the best developed survey capacities in the world and a long tradition of collecting data on a range of economic indicators (Deaton & Kozel, 2005). Although Indian consumption data have been subjected to intense scrutiny, agricultural statistics have eluded the attention of researchers, especially data on crop area statistics. Information on crop area and land use, however, is vital for effective policy planning and designing interventions to fully realize agriculture's potential strengths. In this paper, we extend this literature by drawing on the extensive deployment of geospatial technology in the Indian state of Karnataka to collect crop area statistics in parallel with contemporary data collection methods, thus permitting comparison of the crop area estimates obtained by the two methods.

The objectives of this paper are threefold. First, we document the traditional method for collecting agricultural statistics in India. Second, we develop an alternative data collection method by integrating a geographic information system (GIS) with a global positioning system (GPS) to enhance data quality. Third, we compare the data obtained by the traditional and new methods to assess how well the two measurement methods agree.

Methods and Data

Data

The geospatial crop area survey for this study using GIS/GPS technology was carried out in partnership with the specialized geospatial company Zoomin Softech. Zoomin Softech assisted us in gathering and storing crop information for approximately 2,700 acres of land in the Indian state of Karnataka. This is a typical region located in the Nallur *gram panchayat* (GP) of the Gubbi taluk in the Tumkur district.[1] This region has a mix of irrigated and dry crops, land holdings of various sizes and a diverse occupational structure. A detailed map with survey numbers of each plot of land, along with other maps of the Karnataka State, is presented in Figure 1. Apart from mapping the crop area, the survey also included fallow land, scrub land, water streams, roads, water tanks/ponds and habitation.

Figure 1 here

A large-scale print of the regional map and the land register for the Nallur GP was obtained from the Department of Planning of Karnataka State. The crop inventory, available from the *Pahani book*, or the record of Rights, Tenancy and Crops (RTC)[2], of January 2011 was also collected. The owner of each crop plot area was asked to show and walk along the boundary of his/her land. The field crew also walked along the boundary of the parcel with the GPS device. When the traverse was closed, the details were recorded, and the crop grown

[1] GP is the smallest local government unit in rural areas in India, comprising 3–5 villages with a total population of approximately 5,000. A Taluk comprises several GP's (generally 30–40 GP's, more or less depending on the size of the Taluk) and is a subdivision of a revenue district, which in turn is a subdivision of a state.

[2] *Pahani* (RTC) is a book that lists attributes of land holdings, irrigation, properties, crop types and areas developed under the *Bhoomi* project. Bhoomi is a project for on-line delivery and management of land records in Karnataka.

was identified. The source of the water supply for irrigation was also noted, and the structure, if any (i.e., bore well/open well/canal), was recorded using the GPS device. The crop areas in acres thus obtained under the alternative methods were first collected in January 2011 and again in November 2011. The crop area maps for both these months are presented in Figure 2.

Figure 2 here

Using the GIS application developed by Zoomin Softech, information for each parcel of land was populated with information on the land ownership, the crop area, the crop variety grown, the irrigation facility and the survey number. The field notes used by Zoomin Softech were used in the design, development and implementation of the geospatial survey. These field data were corroborated and supplemented with information collected in interviews with the village accountant.

Current approach and challenges

Under the current scheme, the collection of crop area statistics is assigned to the village level government functionary known as the *patwari* or the village accountant, who is expected to provide timely information using the conventional method. This involves manually gathering data about each crop in each village.[3] Traditionally, the village accountant (VA) is the person responsible for gathering all of the crop information. Approximately 4,600 acres of land in a GP is allocated to each VA for the collection of crop information. To corroborate and systemically document the conventional method, we carried out detailed interviews with two VA's from two different GP's in the Gubbi Taluk—Nallur

[3] A Village Accountant is the administrative head of a "revenue circle", the lowest unit in the revenue administration hierarchy. A revenue circle has 3–4 villages on average, covering approximately 3,800 acres on average. Some *gram panchayat* have two revenue circles and some revenue circles fall in two or more *gram panchayats*.

and Marashetty Halli, chosen to adequately represent the spatial diversity of the data collection method.[4] Both interviews with the VA's were recorded using a voice recorder with prior permission from the respondents; however, the names and locations of the respondents are kept anonymous here for ethical reasons. A flow chart illustrating the conventional method is presented in Figure 3.

Figure 3 here

Each VA is assigned to collect crop information from the 4,600 acres allocated to him in all three seasons of the year. The VA goes to the crop area, visually maps the crop area and enters all the relevant details into the *pahani* book (Bhoomi 2012). *Pahani* contains details of land ownership, area measurement, soil type, nature of possession, liabilities, tenancy and crops grown. The VA is required to use one book for five years to store the details. These registered data are usually verified by the Revenue Inspector (RI) using previous year's crop area data. In the case of no corrections, the data are sent back to the VA for further processing. The VA sends the verified data to the computer center (CC), which in turn sends the data to a private software firm for digitization. The private software firm takes approximately 20 to 30 days to digitalize the data and record it on a CD. The CD is returned to the CC for data uploading into an online database called *Bhoomi*.

Realistically, considering the VA's work load, his ability to collect crop information is limited to half of the total allocated area at most. Moreover, one month allocated to complete the data collection process in each season seems inadequate. Consequently, the major drawback of the conventional method is the lack of quality information on crops grown. The crop area observed from the RTC for the current season and the yield information, gathered

[4] A copy of the questionnaire can be requested from the corresponding author.

from samples collected during crop-cutting experiments in the previous season, are used to estimate the production of crops in the forthcoming season and hence to predict crop prices. Inaccurate crop area statistics thus have a direct bearing on the predicted prices, resulting in flawed policy making and erroneous procurement processes (India's paradox of hunger amidst plenty) and inadequate preparedness to address fluctuating production, which also directly affects farmers to a significant extent.

Appropriateness of the alternative methods

This section evaluates two available geospatial technologies that could be employed to improve the quality of crop area statistics and address the problems with the conventional method described above.

Satellite remote sensing

Remote sensing (RS) is a potential approach to crop area data collection, crop area assessment and forecasting. It provides multi-spectral, synoptic and repetitive coverage with less scope for human intervention in the data generation process, reducing non-sampling errors. This method can be used for anomaly detection amid high temporal resolution, with at least 5–6 observations per season (Ray, Panigrahy & Parihar, 2008). RS techniques gather crop area information when the crop has grown sufficiently (Srivastava, 2011). It can correlate physical properties of soil, such as soil water content, organic matter content, and soil texture, to spectral reflectance. It is also capable of integrating biophysical parameters (such as temperature and leaf area index). This method takes approximately 24–48 hrs to acquire, correct and process the data. However, the time to process a given area depends on the resolution, 1-m-resolution data taking more time to cover an area than 60-m-resolution data. The resolution depends on the type of satellite used.

Although this method has been widely used elsewhere, the Government of India (GOI) only adopted this method with the launch of the program for Crop Acreage and Production Estimation (CAPE) in 1987, covering all the major cereals, pulses and oilseeds produced in India. Following huge losses in 1998 due to a late decision about wheat imports, in August 2006, this program was further strengthened with the commencement of forecasting agricultural output using space agro-meteorology and land-based observation (FASAL). FASAL provides in-season multiple forecasts using weather data, economic factors and land based observations and is capable of producing multiple crop forecasts, starting from sowing to the end of the season (Parihar & Oza, 2006). It also has the potential to forecast changes in cropping patterns, soil moisture and rainfall. Key crops covered under the FASAL are rice, wheat, cotton, sugarcane, rapeseed/mustard, rabi-sorghum, winter potato and jute.

The satellite image associated with this method, however, has a major drawback of not being enlarged beyond 1:10000 (Tsiligirides, 1997). Timely and reliable crop estimates cannot be given for areas with persistent cloud cover, which blocks the satellite view. However, usage of synthetic aperture radar (SAR) can identify the crop even during periods of cloud cover. Integration of optical and SAR images also increase the accuracy of crop mapping (McNairn, Champagne, Shang, Holmstrom & Reichert, 2009). Furthermore, the accuracy of crop inventory data collection using this method can be further improved when combined with field surveys (Mehta, 2000). However, this method appears inappropriate for application to India because of the heterogeneous nature of the cropping patterns and small plot sizes (Ray, Panigrahy & Parihar, 2008).

Geographical information systems and tools

The second geospatial technology considered here is an integrated approach involving both a geographic information system (GIS) and a global positioning system (GPS). Because GIS and GPS technologies are more adaptable and easier to use than RS (Nelson, Orum & Jaime-Garcia, 1999), they have been chosen as the alternative approach for this study. Also, due to the existence of small crop sizes and mixed crops in India, GIS/GPS based systems are better suited for application to India than RS. Previous instances of successful experimentation with this technology have been documented elsewhere (Reichardt, Jurgens, Kloble, Huter & Moser, 2009; ESRI, 2008; Dwolatzky, Trengove, Struthers, McIntyre & Martinson 2006; Murray & Tong, 2009).

Figure 4 here

A flow chart illustrating this alternative method is presented in Figure 4. With this method, an important process improvement is made to ensure easy recording of data in subsequent rounds of crop area recording. Instead of traversing each crop area, each farmer-demarcated parcel within each survey number[5] is traversed using a GPS device in the company of the owner to record the boundary. Farmer-demarcated parcels are small sub-plots within a survey number, created based on topography to identify single crops. From season to season and year to year, the crop in a parcel may change, but changes in a parcel boundary are rare. This provides a detailed base map for crop area data collection. For the geospatial application to provide accurate results, it is recommended that the first survey be implemented rigorously by traversing every single farmer-demarcated parcel of land within a survey number. Corresponding irrigation facilities can also be documented using a GPS device. If a single parcel or sub-parcel of land has more than one crop, the boundary of each

[5] A survey number is an officially demarcated and recorded plot with a specific identification number. A survey number may have multiple owners and crops.

crop plot needs to be traced using the GPS device to record the details of each crop. To improve the accuracy of the data, mapping of the entire geographic terrain within the village is recommended, including all the survey numbers, fallow land, scrub land, water streams, roads and water tanks and pond. This map is then superimposed on the cadastral map for authentication. The data from the GPS device is uploaded to a server through the internet whenever possible.

For this study, Zoomin Softech developed the application and designed the knowledge data base using RTC records and village area maps. The GPS device used was the HTC WildfireS with built-in memory of 512 Mb, an SQLite database and a 5-Mbps camera. The positional accuracy of this device is approximately 2–3 meters, which was calibrated for consistency more than 500 times for a location and later deployed for field application. A seamless geographic database documenting the disposition of the land was also developed, containing village, GP, taluk and district boundaries and the locations of village settlements. The spatial database of the village of Nallur and the web-based GIS application were developed using ArcInfo GIS and ArcGIS Server, respectively. For the development of the GIS server application, the asp.net application framework, the c# and JAVA script programming languages and SQL Server 2008 RDBMS (relational database management system) were used. Specific applications were developed using these software tools to capture (trace) the boundaries of plots and calculate crop areas corresponding to photographs of the crops. These data were then transmitted and synchronized with the GIS database in the server. A screenshot of the software developed for different crops in Nellur is presented in Figure 5. We also present screenshots for horticultural and cereal crops in Figure 6 and Figure 7, respectively. The GIS application developed by Zoomin Softech updates the changes in the server and functions as a graphical user interface (GUI) for the user. One advantage of this method is that it is sufficient to update and map only those crop areas that are subject to seasonal flux, keeping the operational cost of the data collection lower in subsequent rounds.

Figure 5 here

Figure 6 here

Figure 7 here

Comparing alternative approaches

The information gathered by interviewing the VA's were transcribed and interpreted to identify the processes involved in the conventional method, which was then compared to the alternative method proposed in this paper. The key differences between the two methods are briefly described here and documented in detail in Table 1. The differences between the processes involved in these approaches can be classified into three categories: (a) data collection, (b) data verification and (c) data digitization and dissemination.

Table 1 here

Process of data collection: In the alternative approach, data collection is completely digitized, reducing the time for collection and dissemination of information. Under the conventional method, crop area data is gathered by visual observation and recorded manually in the *Pahani* book. Under the alternative method, crop area data are gathered and recorded using a GPS device while traversing the parcel with the farmer and then digitally transferring the information to the database. The GPS locations traversed during November 2011 are presented in Figure 8. The automated process in the alternative method helps ensure the accuracy of the data. Adequate recording of the corresponding irrigation facilities is also possible using the alternative method.

Figure 8 here

Verification of data: The data collected using the conventional method is verified by the Revenue Inspector (RI) using previous RTC records. In the case of the alternative method, the data are verified by comparing the digitized RTC records with the owners of the crop areas while traversing the parcels.

Digitization and dissemination of data: The digitization of the crop area gathered by the VA using the conventional approach takes approximately 20–30 days. However, using the alternative approach, the data collected are digitized using a GPS device, and the data are uploaded to the server instantaneously through the internet. Another drawback of the conventional approach is the lack of a GUI to display crop area information. The GIS application gives micro details of the crop area data, facilitating accurate forecasting of crop areas.

The Bland–Altman method

The Bland–Altman approach (Bland & Altman, 1983; Bland & Altman, 2012) employed here to test the agreement between the two methods of measurement can be represented as follows:

$$y_{mi} = \alpha_m + \mu_i + e_{mi} \quad e_{mi} \sim N(0, \sigma_m^2)$$

where y_{mi} is the measurement by method m for land parcel i, α_m is the crop area averaged over individual methods, μ_i is the crop area averaged over land parcel i, and e_{mi} is the disturbance, with zero mean and constant variance. Here m signifies the two methods of measuring crop area: (i) the conventional method c and (ii) the alternative method a. The difference in measurement between the methods is d, where values of $d_i = y_{ci} - y_{ai}$ are identically distributed with mean $\alpha_c - \alpha_a$ and variance $\sigma_c^2 + \sigma_a^2$ and are independent of the averages $\bar{y}_{.i}$ if $\sigma_c = \sigma_a$ or r = 0. Here, r is the correlation between the mean and the variance. The Bland–Altman plot of d_i versus $\bar{y}_{.i}$ is used to inspect visually whether the difference and its variance are constant as a function of the average. From this plot, it is much easier to assess the magnitude of the disagreement between the methods, spot outliers, and determine whether there is any trend. If the measurements from both methods are comparable (agree), the differences should be small and centered around zero and should exhibit no systematic variation with the mean of the measurement pairs.

Results and discussion

This section compares the crop area data collected using the alternative method (GIS/GPS technologies) with the administrative data collected by the conventional method (RTC records) described previously.

Comparing methods of measurement

As Table 2 illustrates, a total of 19 crops are grown, only 13 of which are listed in the RTC, covering, on average, only 42% of the total number of plots monitored using the conventional method. Note that the crop area information for both methods presented in this table is not based on matching the plot-wise information collected using each method. A comparison of column 3 and column 6 in Table 2 shows that the conventional method covers only 63% and 47% of the total number of plots for coconut and finger millet, respectively. The worst coverage is for arecanut (10%) and mango (34%). As is typical in agriculture in many developing countries, most farmers cultivate a mix of both subsistence and cash crops, with a portfolio of short-duration and long-duration crops (Patarasuk & Binford, 2012). The remainder of the analysis presented below is based on plot-wise matching of the crop area data reported using the two methods.

Table 2 here

Although the overall difference in the total crop area estimates from the two methods is 56%, the discrepancies depend on the types of crops grown. The differences in crop area estimated for each crop using the conventional and alternative methods are presented in Figure 9. The crop areas in acres obtained using the alternative method are the simple averages of the estimates obtained twice, first in January and again in November 2011. The differences, reported here in acres, are measured for each crop along the ray from the center. The differences are negligible for some crops, such as groundnut, eucalyptus, chilies, beans, bananas, teak, pepper, flowers, beetle leaf, tamarind, sapodilla and sorghum. However, these constitute insignificant total crop areas of 2.5% and 1.6%, as estimated using the conventional and alternative methods, respectively.

Figure 9 here

Figure 9 shows that the largest absolute difference in crop area estimates between the methods is for finger millet (54%). This short-duration staple crop constitutes approximately 30% of the total crop area. Underreporting of crop areas for short-duration crops have also been reported by Srivastava (2003). For coconut, the underestimates by the conventional method are somewhat lower (27 percent); however, this crop constitutes a larger total crop area of approximately 38%. The other crops for which considerable differences were observed in the estimates obtained using the two methods are arecanut and mango.

Note that except for finger millet, all the other crop area estimates that exhibit considerable differences correspond to high-value, long-duration cash crops. Because these cash crops constitute approximately 63% of the total crop area, it is paramount to investigate the reasons for the discrepancies in the crop area estimates. This is surprising, given that long-duration crop production is easily predictable using the conventional method, as such crops remain planted for several years, while short-duration crop production can potentially vary from one season to the next. However, discussions with farmers pointed to changing cropping patterns as the key reason for the discrepancies observed for long-duration crops. Over the years, the crop area under all three cash crops has expanded, while the area under finger millet has contracted. These changing cropping patterns, not captured and reflected in the administrative data collected using the conventional method, have wider implications for access to crop loans and crop insurance and could also pose serious threats to food security.

A comparison of crop areas estimated using the two methods shows that the conventional method, in general, underestimates crop areas and is not appropriate for capturing changing cropping patterns. This is an enormous concern for a developing country with its agriculture sector in transition to commercialization and adoption of high-value crops.

In the next section, we examine whether the two methods yield significantly different results using the Bland–Altman approach. This approach is used extensively for comparison of methods in the medical and biological sciences (Bland & Altman, 1986; Euser, Finken, Keijzer-Veen, Hille, Wit & Dekker, 2005). Here, the emphasis is on examining whether a low-cost conventional method is comparable to an expensive alternative method to the extent that one might replace either one with the other and ensure sufficient accuracy in measuring the area under each cultivated crop.

The Bland–Altman method is supplemented with a more formal test, Pitman’s test of difference in variance (Pitman, 1939; Snedecor & Cochran, 1967), which compares two correlated variances in paired samples to test the degree of agreement between the conventional and alternative methods for measuring crop areas. The results of this test are reported in Table 3 for all of the crops considered. The comparisons are based on plot-wise matching of crop area information obtained using both methods. The total number of observations under "All Crops" (last row) does not match the total number of observations across crops, due to a mismatch in cultivated crops recorded under both methods across all crops. Apart from the crops listed in this table, "All Crops" also includes bananas, beans, chilies, eucalyptus, groundnut, sapodilla, tamarind, teak and pigeon pea. These crops were excluded from the disaggregated analysis due to insignificant crop areas under each of these categories. The first two columns show the estimated bias with the expected intra-individual differences' 95% confidence interval (CI) limits. The third column shows the mean difference plus or minus 2 standard deviations ($\bar{d}$ ± 2sd). The results of Pitman's test are reported in columns 4 and 5, with the correlation between the difference in the methods and their average denoted by r. The next column reports the p-value of a test of the null hypothesis that there is no significant difference in the variances between the conventional and alternative methods.

Table 3 here

Figure 10 here

The Bland–Altman plot for the total crop area ("All Crops") presented in Figure 10 illustrates the presence of outliers and the existence of an association between the differences and the sizes of the measurements. A log transformation did not alter the results to any great extent. The plot displays considerable lack of agreement between the conventional and alternative methods, with discrepancies stretching the limits of agreement (-2.1 and 2.9) beyond acceptable levels (Table 3, column 3). The limits of agreement are not small enough for us to be confident that the conventional method can be used in place of the alternative method. The results of the test of independence (null hypothesis of r = 0), presented in Table

3, columns 4 and 5, show that a significant relationship exists between the difference between the methods and the size of measurement ($r = 0.21$, $p = 0.00$). These results confirm the lack of agreement between the methods for "All Crops" (the last row in Table 3).

Similar results were also obtained for all the long-duration, high-value crops: arecanut, coconut and mango. The bias, indicated by the mean difference in Table 3, column 1, is largest for arecanut at 0.81, while a lower r ($r = 0.12$, $p < 0.10$) is observed for coconut (Table 3, column 4 and column 5). However, the difference is significant only at the 10% level. For all three high-value crops, the mean differences indicate a significant bias in crop area estimation by the conventional method (observed in Figures 11, 12 and 14 and confirmed in Table 3, column 1) and the lack of agreement between the methods (Table 3, column 4 and column 5).

Figure 11 here

Figure 12 here

Figures 13 here

Figures 14 here

Figures 15 here

Figures 16 here

As noted in the previous section, the results for the short-duration staple crops, sorghum, paddy and finger millet, reported in Figures 13, 15 and 16, respectively and in Table 3 are somewhat surprising. The mean difference of 0.05 for paddy reported in Figure 15 and in Table 3, column 1, shows negligible bias. The mean differences for sorghum and finger millet (Table 3, column 1), however, exceed acceptable levels, indicating underestimation of the crop area by the conventional method in comparison to the alternative method. However, Pitman's test detected no significant difference between the variances associated with the conventional and alternative methods for any of the three crops (Table 3, column 4s and 5). We therefore cannot reject the null hypothesis of no correlation between the difference between the methods and the size of the measurements and we conclude that good agreement exists between the two methods.

The analysis described above yields some interesting results. First, the conventional method, which entails manually gathering data, does not capture changing cropping patterns stimulated by commercializing agriculture in a developing country (Patarasuk & Binford, 2012). Comparisons of the estimates of areas under crops and the types and number of crops reveal considerable discrepancies between the two methods. The conventional method provides information for only 13 of the 19 crops grown in the study area and ignores some of the vital high-value cash crops in transitional agriculture. The crop area estimates obtained using the alternative method differ significantly from the estimates obtained using the conventional method (by 56%), suggesting that routinely collected administrative data on crop areas are likely to underestimate actual crop areas. This underestimation could significantly affect projections of crop production and underestimate actual production. The resulting excess production, with no planning for utilization in place, could well result in rotting food stocks, which is a recurring problem in India (Basu, 2010). Second, the conventional method seems appropriate for measuring areas under crops with minimal year-to-year changes, such as cereals, but not for high-value cash crops. The discrepancies in the area estimates between the two methods for some cash crops are over 80% (84% for arecanut and 96% for tamarind, for instance). Changes in the magnitude and direction of these differences across crops can be useful in identifying ways to improve the quality of area statistics. Third, although the initial application of geospatial tools may not be cost effective, the cost of subsequent updating is lower than that of the conventional method.

Conclusions

Despite the significance of agriculture in developing countries and the general recognition of the need for improving agriculture and rural statistics in these countries, surprisingly little research on this topic exists. This paper contributes to this literature by focusing on how agricultural statistics can be strengthened in developing countries using new geospatial tools taking the case of rural Karnataka in India. We implemented a comprehensive survey of crop area using the GPS/GIS tools in parallel with the conventional method to document any differences between the recorded crop area estimates for the same plots of land.

India has a long tradition of generating comprehensive crop area and land use statistics using decentralized village-level agencies, with little systematic evaluation of the data generated. However, new technologies offer the potential to improve measurement accuracy by rigorously evaluating data, considered by many Indian policy makers to be no more reliable than folk wisdom. The results presented here suggest that the conventional method does not seem to capture changing cropping patterns resulting from the commercializing of agriculture in developing countries; however, it seems to be appropriate for measuring crop areas under staple crops, although not for high-value cash crops. The major reason for the poor quality of crop area and land use statistics is the inability of village accountants to devote adequate time and attention to data collection using the conventional method. Hence, policies aimed at strengthening and modernizing this legacy of the Indian data system with new geospatial tools can potentially contribute to strengthening food security, augmenting agricultural price policies and improving predictions obtained from crop and land use models. As the accuracy of the estimates of food production are primarily dependent on the accuracy of crop acreage estimates, the new approach would help in generating more accurate data on food production.

Although this paper demonstrates the merits of using geospatial technology in collecting crop area information, there are also potential payoffs in routinely deploying this technology for household surveys, household asset and resource mapping, geo-referencing of village infrastructure and geo-referenced poverty mapping. Some recent studies highlight the efficacy of GIS applications in poverty reduction (Baker & Grosh, 1994; Lang, Barrett, & Naschold, 2012). With extensive parcel mapping, it is possible to develop a self-reporting-based crop area system in which each parcel mapped is given an identification number and farmers are made aware of these identification numbers. Farmers can then report the crops they intend to grow or are growing, either in person or over the phone, to a public or private agency in charge of the crop area database. The agency can do sample checking of the farmers' reported data through field visits, using hand-held GPS devices. With falling costs and increasing evidence of the potential benefits, this technology will see wider applications within developing countries.

The amount of credit a farmer can obtain from formal financial institutions depends on the area under each crop and the estimated cost of cultivation. Cash crops generally require larger amounts of credit than food crops. Under-recording of cash crop areas in RTC will limit the availability of credit to farmers. Similarly, insurance coverage and premiums to be paid are specific to crops. Inaccurate recording of crop areas makes it difficult to offer crop insurance to farmers; hence, insurance providers mainly focus on farmers who have previously received credit from formal institutions, such as banks and cooperative societies. Offering crop insurance to farmers who have not obtained credit from other sources involves additional costs to the agency for physical verification of the crop area, and therefore, these farmers are not generally covered by the insurance companies. Ironically, such farmers are typically the small and marginal farmers who need insurance most.

Some analytical caveats remain, however. First, the results presented in this paper are specific to the survey region in the Indian state of Karnataka, although the implications and issues raised are highly relevant to the rest of India, where the conventional method is still widely employed in gathering crop area statistics. A second critique applies to the usage of GIS/GPS technology, which requires manually traversing the crop area accompanied by the crop owner. An unscrupulous crop assessor could choose to ignore the directions of the crop owner. This geospatial survey was subjected to strict quality controls requiring the presence of the crop owner and was also independently monitored by a supervisor. This was a comprehensive survey that accounted for each parcel of land within each survey number in the survey region.

Third, more generally, GIS/GPS technology is not a panacea and neither are other relatively new technologies, because the success of the technology also depends on proper use, proper data management and an appropriate data transfer system. The specific geospatial survey developed by Zoomin Softech and used in this study required considerable resources to refine the application, based on the inputs from the RTC records and village area maps, and design the knowledge base. For the geospatial survey to be robust, this technology requires traversing every plot of land within the survey region in the first survey. Hence, budget considerations may limit the use of this technology. However, with time, the cost of the technology may fall, enabling wider use of this technology and improving the accuracy of a range of statistics.

References

African Development Bank Group (2011). *Improving Statistics for Food Security, Sustainable Agriculture, and Rural Development: An Action Plan for Africa (2011-2015)*, African Development Bank Group, Ghana. Food and Agriculture Organization (1999) Crop Information Systems. http://www.fao.org/sd/EIdirect/EIre0078.htm. Accessed 31 August 2012.

Baker, J. & Grosh, M. (1994). Poverty reduction through geographic targeting: How well does it work?, *World Development*, 22(7), 983-995.

Basu, K. (2010). *The Economics of Foodgrain Management in India*, Ministry of Finance, Government of India, New Delhi.

Bell, K. & Dalton, T. (2007) Spatial economics analysis in data-rich environments, Journal *of Agricultural Economics*, 58(3), 487-501.

Beegle, K., Carletto, C., & Himelein¸ K. (2012). Reliability of recall in agricultural data, *Journal of Development Economics*, 98(1), 34-41.

Bhoomi. (2012). *Pahani (RTC)*. Retrieved 02 17, 2012, from Bhoomi: http://bhoomi.karnataka.gov.in/faq.htm#What is Pahani(RTC)?

Bland, J. & Altman, D. (1983). Measurement in medicine: The analysis of method comparison studies, *Statistician*, 32(3), 307-317.

Bland, J. & Altman, D. (1986). Statistical methods for assessing agreement between two methods of clinical measurement, *Lancet*, 1(8476), 307-310.

Bland, J. & Altman, D. (2012). Measurement method comparison, *Anesthesiology*, 116(1), 182-185.

Caeyers, B., Chalmers, N., & Weerdt, J. (2012). Improving consumption measurement and other survey data through CAPI: evidence from a randomized experiment, *Journal of Development Economics*, 98(1), 19-33.

Deaton, A. & Kozel, V. (2005). Data and dogma: The great Indian poverty debate, *World Bank Research Observer*, 20(2), 177-199.

Deininger, K., Carletto, C., Savastano, S., & Muwonge, J. (2012). Can diaries help in improving agricultural production statistics? Evidence from Uganda, *Journal of Development Economics*, 98(1), 42-50.

Dewan, A. & Yamaguchi, Y. (2009). Land use and land cover change in Greater Dhaka, Bangladesh: Using remote sensing to promote sustainable urbanization, *Applied Geography*, 29(3), 390-401.

Dwolatzky, B., Trengove, E., Struthers, H., McIntyre, J. & Martinson, N. (2006). Linking the Global Positioning System (GPS) to a Personal Digital Assistant (PDA) to support tuberculosis control in South Africa: A pilot study, *International Journal of Health Geographics*, 5(34), 1-6.

Euser A., Finken M., Keijzer-Veen M., Hille E., Wit J., & Dekker, F. (2005). Associations between prenatal and infancy weight gain and BMI, fat mass, and fat distribution in young adulthood: a prospective cohort study in males and females born very preterm, *American Journal of Clinical Nutrition*, 81(2), 480- 487.

ESRI. (2008). *GIS for Sustainable Agriculture.* Retrieved 20-03-2012, from ESRI: www.esri.com/library/bestpractices/sustainable-agriculture.pdf

Kulshrestha, A. & Kar, A. (2005). Consumer expenditure from the National Accounts and National Sample Survey. In Deaton, A., Kozel, V., *The Great Indian Poverty Debate*. New Delhi: Macmillan.

Lang, C., Barrett, C. & Naschold, F. (2012). Targeting maps: An asset-based approach to geographic targeting, *World Development*, 2012.06.006.

McNairn, H., Champagne, C., Shang, J., Holmstrom, D., & Reichert, G. (2009). Integration of optical and Synthetic Aperture Radar (SAR) imagery for delivering operational annual crop inventories, *ISPRS Journal of Photogrammetry and Remote Sensing*, 64(5), 434-449.

Mehta, R. (2000). Remote sensing applications in agriculutral statistics system: Scope, potential and limitations, *Indian Journal of Agricultural Economics*, 55(2), 87-95.

Miller, H. (2010). The data avalanche is here: Shouldn't we be digging?, *Journal of Regional Science*, 50 (1), 181-201.

Mills, J., Curtis, A., Kennedy, B. & Kennedy, S. (2010). Geospatial video for field data collection, *Appled Geography*, 30 (4), 533-547.

Muller, R., Muller, D., Schierhorn, F. & Gerold, G. (2011). Spatiotemporal modeling of the expansion of mechanized agriculture, *Applied Geography*, 31(2), 631-640.

Murray, A. (2010). Quantitative geography, *Journal of Regional Science*, 50(1), 143-163.

Murray, A. & Tong, D. (2009) GIS and spatial analysis in the media, *Applied Geography*, 29 (2), 250-259.

Nelson, M. R., Orum, T. V., & Jaime-Garcia, R. (1999). *Applications of Geographic Information Systems and Geostatistics in Plant Disease Epidemiology and Management.* Arizona: The American Phytopathological Society.

Parihar, J. & Oza, M. (2006). FASAL: an integrated approach for crop assessment and production forecasting. In Kuligowski, R., Parihar, J., Saito G., *Agriculture and Hydrology Applications of Remote Sensing*. SPIE Press.

Patarasuk, R. & Binford, M. (2012). Longitudinal analysis of the road network development and land-cover change in Lop Buri province, Thailand, 1989-2006, *Applied Geography*, 32(2), 228-239.

Pitman, E. (1939). A note on normal correlation, *Biometrika*, 31(1/2), 9-12.

Ray, S., Panigrahy, S., & Parihar, J. (2008). Precision farming in Indian Context - Role of Remote Sensing. In Anandaraja, N., Chandrakandan, K., Ramasubramaniam, M., *Extension of Technologies from Labs to Farms* (pp. 491). New Delhi: New India Publishing Agency.

Reichardt, M., Jurgens, C., Kloble, U., Huter, J., & Moser, K. (2009). Dissemination of precision farming in Germany: acceptance, adoption, obstacles, knowledge transfer and training activities, *Precision Agriculture*, 10(6), 525-545.

Sen, A. (2000). Estimates of consumer expenditure and its distribution: statistical priorities after the 55th Round, *Economic and Political Weekly*, 35(51), 4499–4518.

Snedecor, G. & Cochran, W. (1967). *Statistical Methods*. 6th ed. Aimes, IA: Iowa State University Press.

Srivastava, A. (2003). Crop area and yield statistics through ICS scheme, *Journal of the Indian Society of Agricultural Statistics*, 56(1), 64-87.

Srivastava, S. (2011). *Case study: Harnessing new tools and techniques for making agricultural statistics more efficient and evidence based support food security policy decisions in India*. Retrieved on 17-09-2012 from http://www.unescap.org/stat/meet/data-use-oct2011/india.pdf

Tsiligirides, T. (1997). Remote sensing as a tool for agricultural statistics: a case study of area frame sampling methodology in Hellas, *Computers and Electronics in Agriculture*, 20(1):45-77.

United Nations (1979). *Improving social statistics in developing countries: Conceptual framework and methods*, Studies in Methods Series F, No. 25, Department of International Economic and Social Affairs Statistical Office, New York.

World Bank (2008). *World Development Report 2008: Agriculture for Development.* Washington, D.C.: The World Bank.

World Bank (2011). *Global strategy to improve agricultural and rural statistics*; Report No. 56719-GLB, Washington, D.C.: The World Bank.

Table 1. Methods comparison

Parameters	Conventional method	Alternative method
Cost per season for the total area of 4600 acres (assigned to each VA)	Costs 538.88 US$ (1 US$ = Rs. 55.67) Cost breakdown: 2 months VA salary= 2 X 269.44 US$	Costs 485.86 US$ Cost breakdown*: cost of updating = 414 US$ (0.09 US$ price paid for traversing per acre X 4600 acres) + 26.94 US$ is the user cost of a hand held device + 44.92 US$ paid for verification of data
Connectivity	The digitized data is available in *Bhoomi* database (*Bhoomi* database is operated by govt.) which can be accessed by all stakeholders	The data is directly transferred to the server which can be accessed using GIS application
Capacity	According to VA, collecting 4600 acres in one month is a tough task. Therefore, VA can only collect 50% of the data in one month	The crop inventor had covered 2700 acres in one month
Adequate	The information collected by VA is used by government since many years. Therefore, it should be adequate. However, the quality of the data has deteriorated in recent year	The information collected by crop inventor is capable of providing adequate information using GPS/GIS
Reliable	The data collected by VA is through eye-balling technique and it is stored manually in *Pahani* books which is later digitized and transferred to *Bhoomi* database	The crop inventor collects the data using GPS device and transfers the data to server using internet
Timely	The time required by VA to collect the data is 30 days. It again takes 20-30 days for digitization	The crop inventor collected accurate data in less number of days then VA. The data collected is in digitized format
Security	The data is collected manually and stored in *Pahani* books which can be subjected to risks. The data is then verified by RI. The data is digitized by a third party (i.e. a private player) and is transferred to *Bhoomi* database	The data collected is not manually stored in records, which reduces human intervention. The crop area is traversed using GPS device. The GPS device transfers the data to a server which is accessed authentically
Better Planning of Government	The collection and dissemination of data takes nearly 60 days. The accuracy is poor and the technique for data collection is not reliable	The collection and dissemination of data occurs on the same day. Data has high accuracy and the technique for data collection is also reliable
Effective Delivery	The delivery of data is instantaneous after digitization. However, the delay in digitization and poor accuracy are some of the the drawbacks	The delivery of data is instantaneous after collecting the data using GPS device. There is no delay in digitization and the accuracy is above 90%. GIS

		application provides various options for viewing the data
Easy Monitoring and Evaluation	The data can be easily monitored and evaluated after the data is uploaded in the *Bhoomi* database	The data is easy for monitoring and evaluation from the beginning of the process (i.e. during data collection using GPS device)
Frequency of data collection	The data is collected by VA once every season and is capable of collecting data during anytime of the year	The data is collected by crop inventor during every season. Additional updating is also possible at anytime of the year, irrespective of the climate

Note: * Further disaggregation of the costs and their justifications can be requested from the corresponding author.

Table 2. Type of Plot Area Utilization

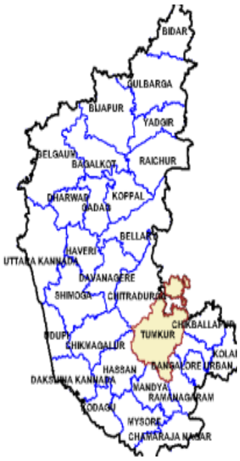
Land use type	Conventional method			Alternative method		
	Mean	SD	N	Mean	SD	N
	1	2	3	4	5	6
Arecanut	1.57	0.91	27	1.04	1.03	250
Banana	0.68	0.37	7	0.43	0.30	8
Beans	2.42	1.12	3	0.66	0.16	3
Beetle Leaf	n.a.	n.a.	n.a.	0.31	0	1
Chilies	0.75	0	1	0.11	0	1
Coconut	1.72	1.43	377	1.51	1.36	591
Eucalyptus	3.07	0	3	1.67	1.44	10
Fallow	0.2	0	4	1.62	1.96	87
Flower-Kakad	n.a.	n.a.	n.a.	0.15	0	1
Government Land	n.a.	n.a.	n.a.	4.29	0	1
Groundnut	0.5	0	1	1.72	0	1
Habitation	0.62	0.89	27	2.11	2.34	5
Horsegram	n.a.	n.a.	n.a.	3.74	3.28	5
Sorghum	1.01	0.87	6	1.35	1.41	8
Mango	1.67	2.05	68	1.51	1.36	202
Paddy	1.11	1.17	15	1.07	1.24	71
Pepper	n.a.	n.a.	n.a.	0.22	0	1
Finger Millet	1.66	1.20	191	1.68	1.67	408
Road	n.a.	n.a.	n.a.	0.44	0.46	17
Sapodilla	0.75	0	1	0.69	0	1

Scrub Land	n.a.	n.a.	n.a.	6.59	6.92	26
Stream	n.a.	n.a.	n.a.	1.12	1.80	5
Tamarind	0.25	0	1	0.56	0.39	10
Tank/Pond	n.a.	n.a.	n.a.	3.53	4.82	8
Teak	n.a.	n.a.	n.a.	1.50	0	1
Pigeon Pea	n.a.	n.a.	n.a.	1.08	1.13	31
All	1.62	1.41	732	1.53	1.80	1753

Note: n.a. refers to information not available in the administrative records (RTC). Mean and SD are calculated from area in acres while N is the number of plots under different land use types. These calculations are not based on plot wise matching (same plot) of crop area information from both methods.

Table 3. Comparison of Methods for Estimating Crop Area

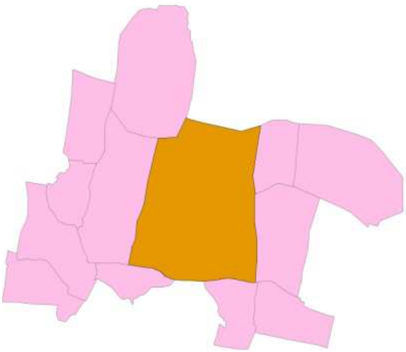
Crop	Mean difference		Limits of agreement	Pitman's test of difference in variance	
	Mean	95% CI		r value	p value
	1	2	3	4	5
Arecanut (n = 148)	0.81	0.51 to 1.10	- 2.83 to 4.45	0.34	0.000
Coconut (n = 458)	0.33	0.21 to 0.45	-2.32 to 2.99	0.12	0.009
Sorghum (n = 11)	0.33	-0.08 to 0.75	-0.91 to 1.57	0.43	0.180
Mango (n = 127)	0.40	0.11 to 0.68	-2.89 to 3.69	0.49	0.000
Paddy (n = 44)	0.05	-0.22 to 0.32	-1.73 to 1.83	0.25	0.089
Finger Millet (n = 249)	0.32	0.18 to 0.45	-1.84 to 2.49	0.00	0.886
All crops (n = 655)	0.36	0.27 to 0.46	-2.16 to 2.90	0.21	0.000



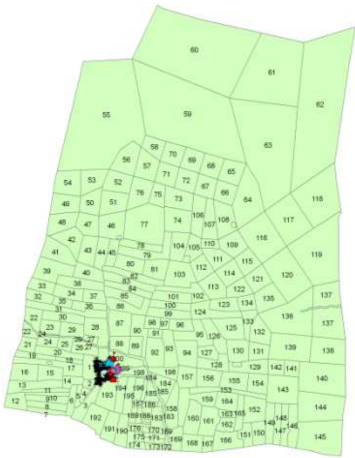
Karnataka State

Gubbi Taluk inTumkur District

Nallur GP in Gubbi Taluk



Nallur Village in Nallur GP



Nallur Village map

Figure 1. Map of the Karnataka State, Gubbi Taluk, Nallur GP and Village

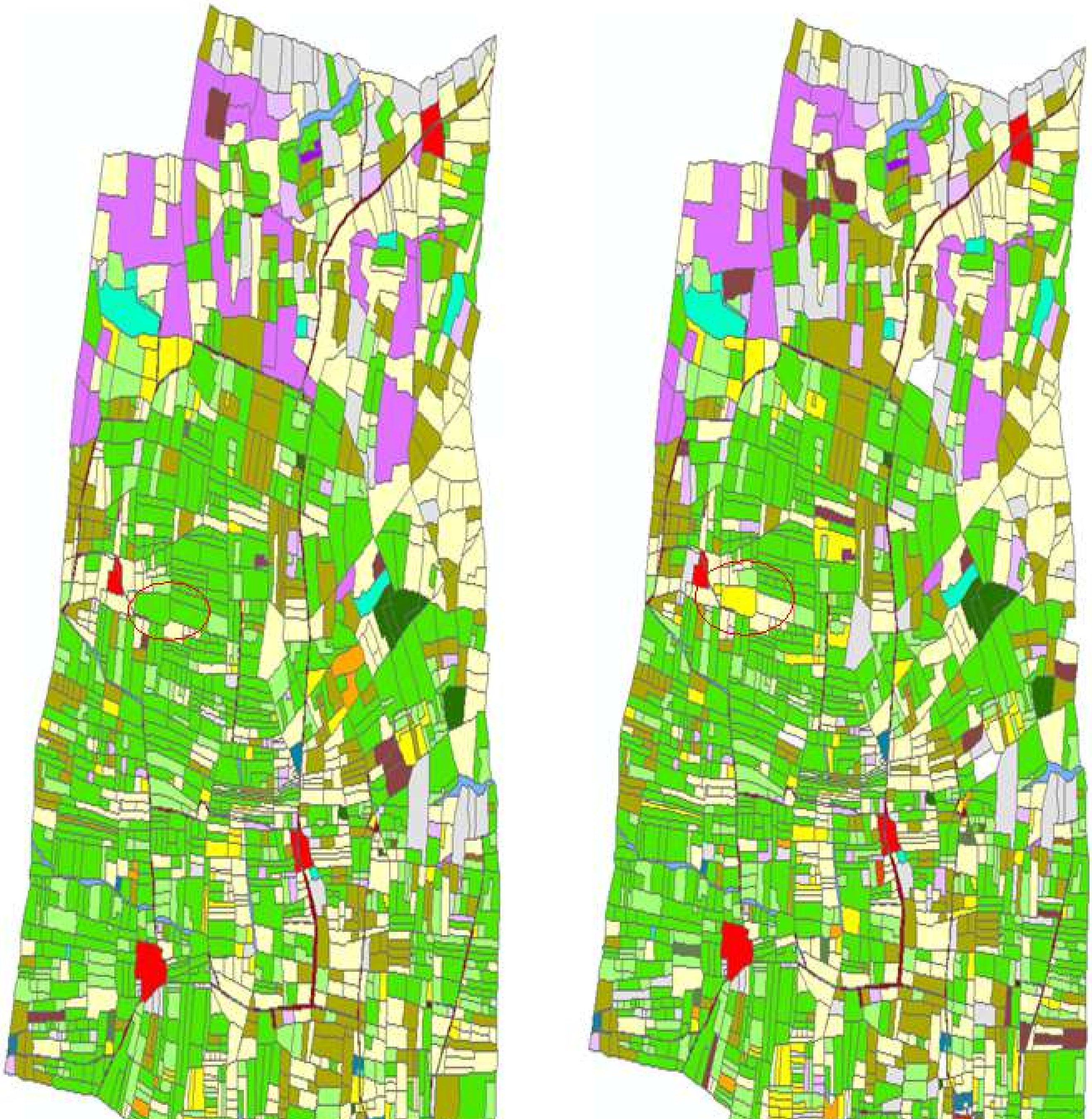


Figure 2. Crop map for both seasons separately

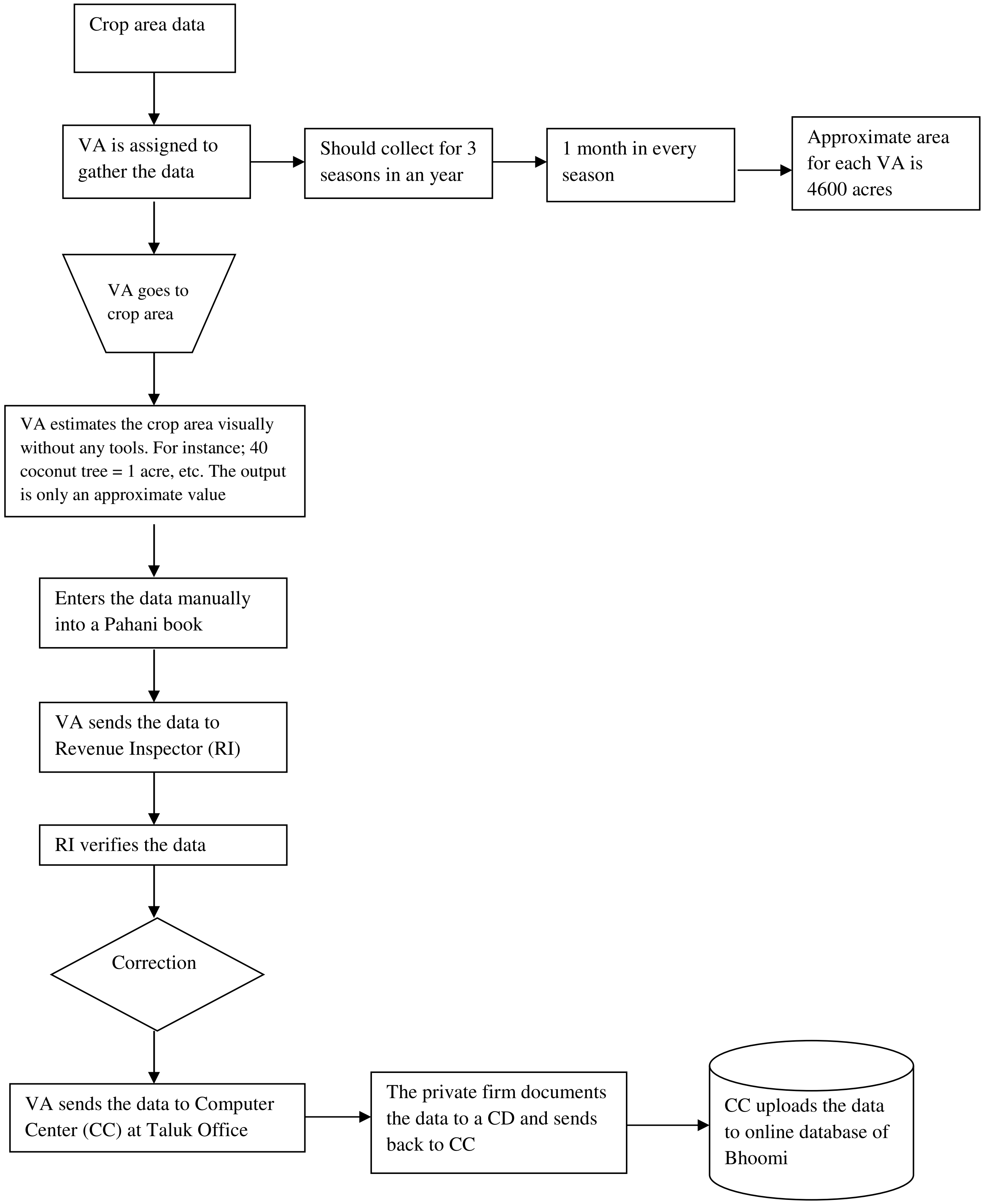
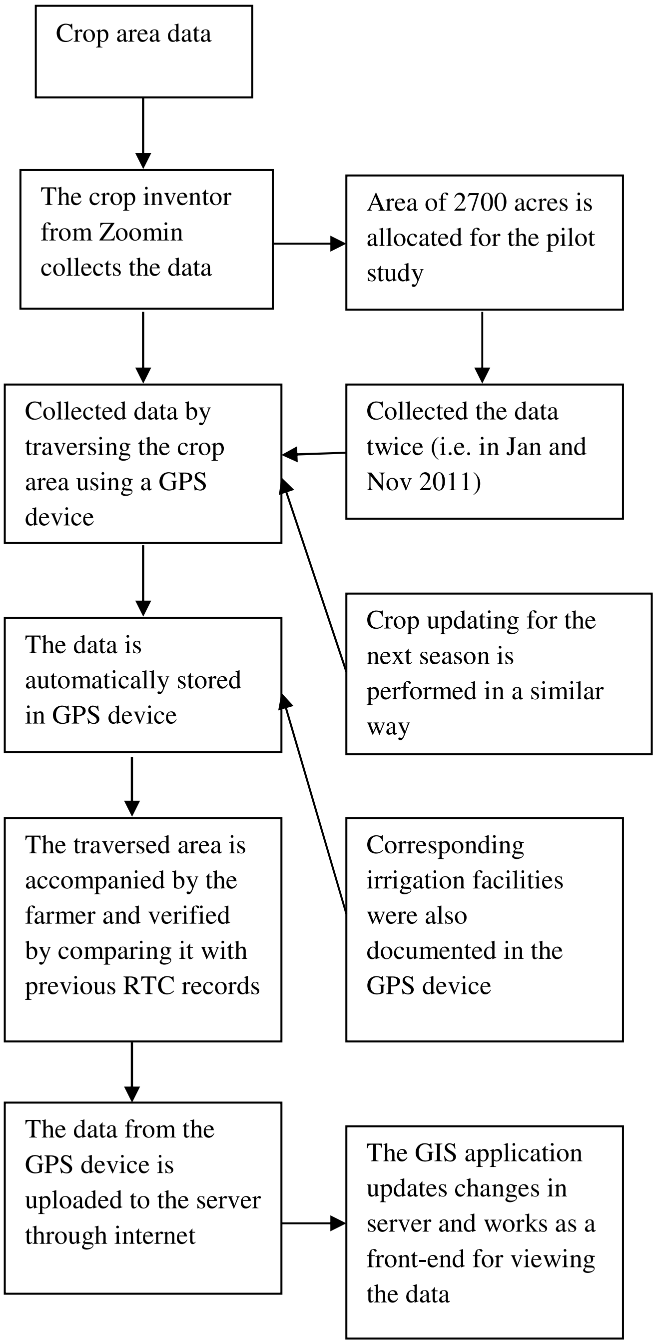


Figure 3. Summary of the conventional method



Since, all the crops are not subjected to change during each season; the crop inventor updates the crops which are subjected to change using previous season's crop area map as a reference

Figure 4. Summary of the alternative method

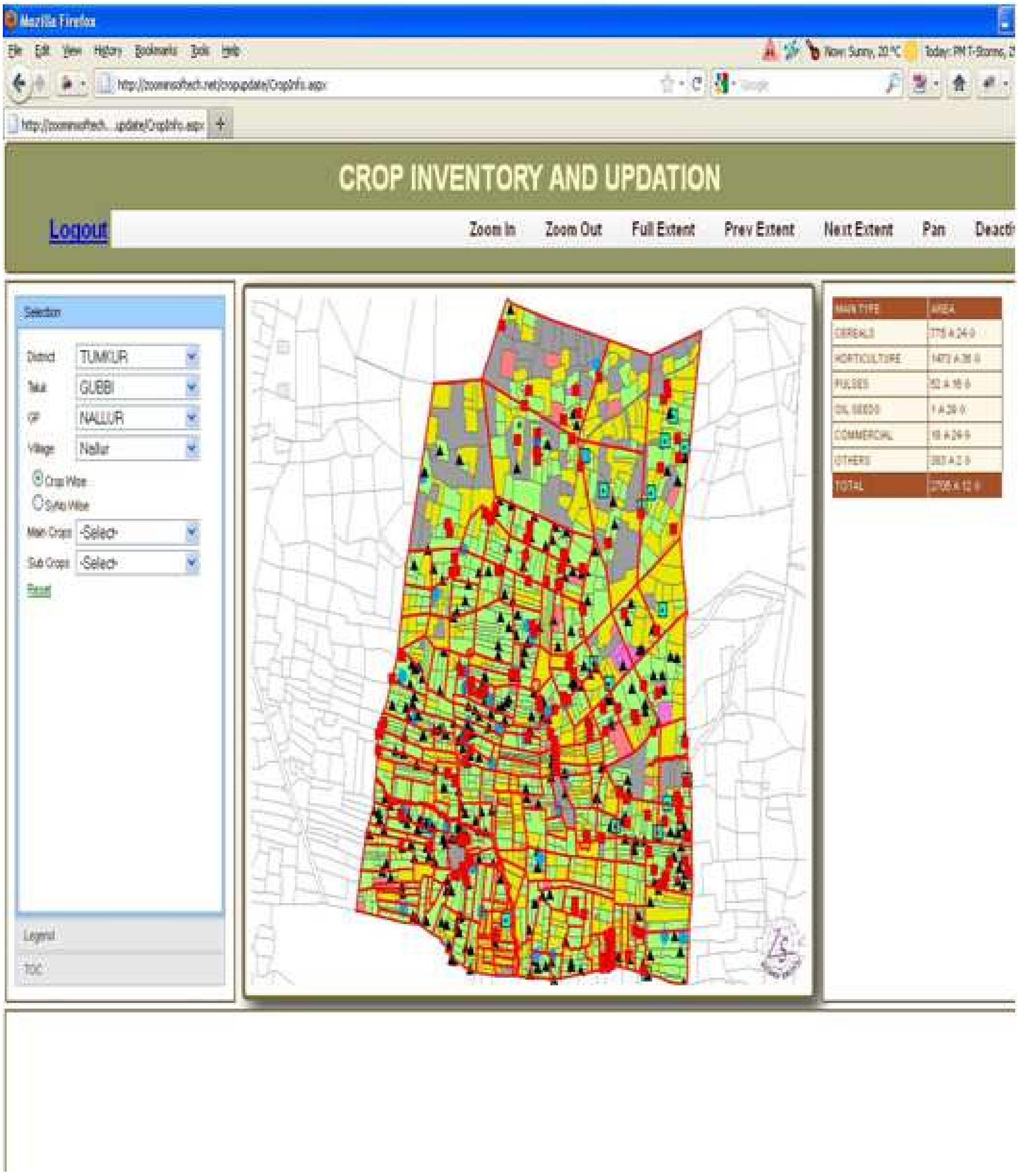


Figure 5. GIS application presenting different crops in Nallur

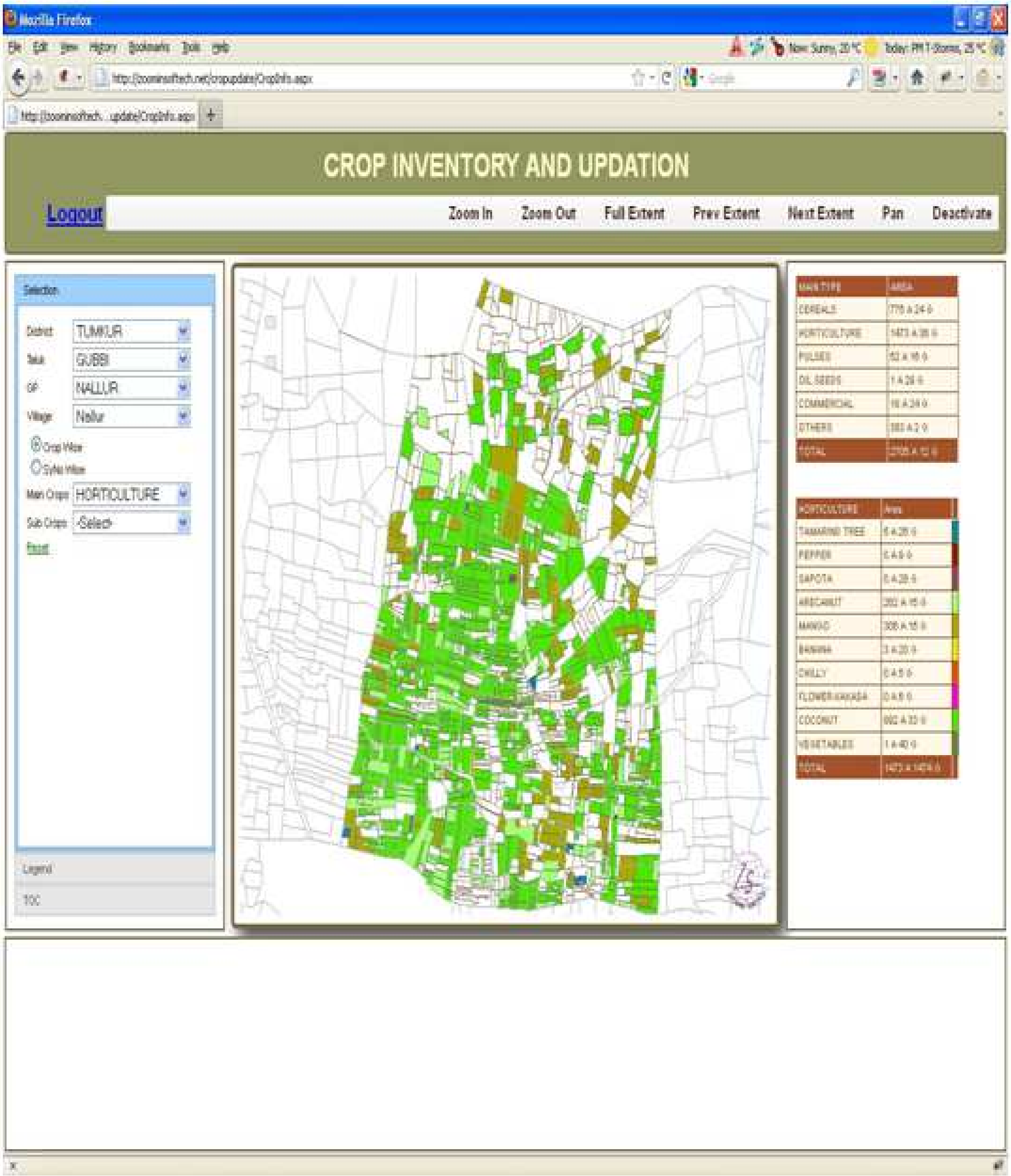


Figure 6. GIS application presenting horticulture crops in Nallur

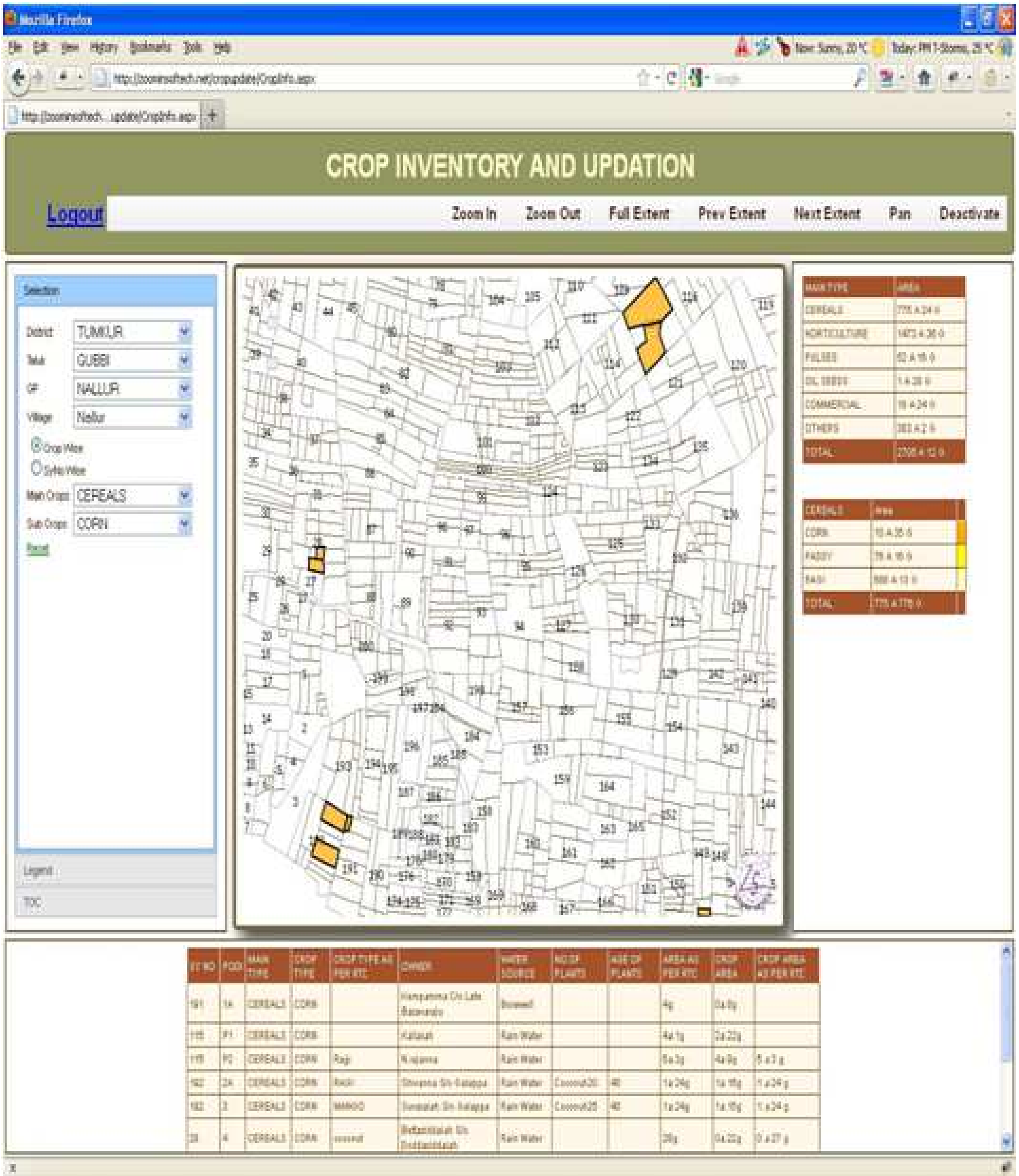


Figure 7. GIS software application presenting cereal crops in Nallur

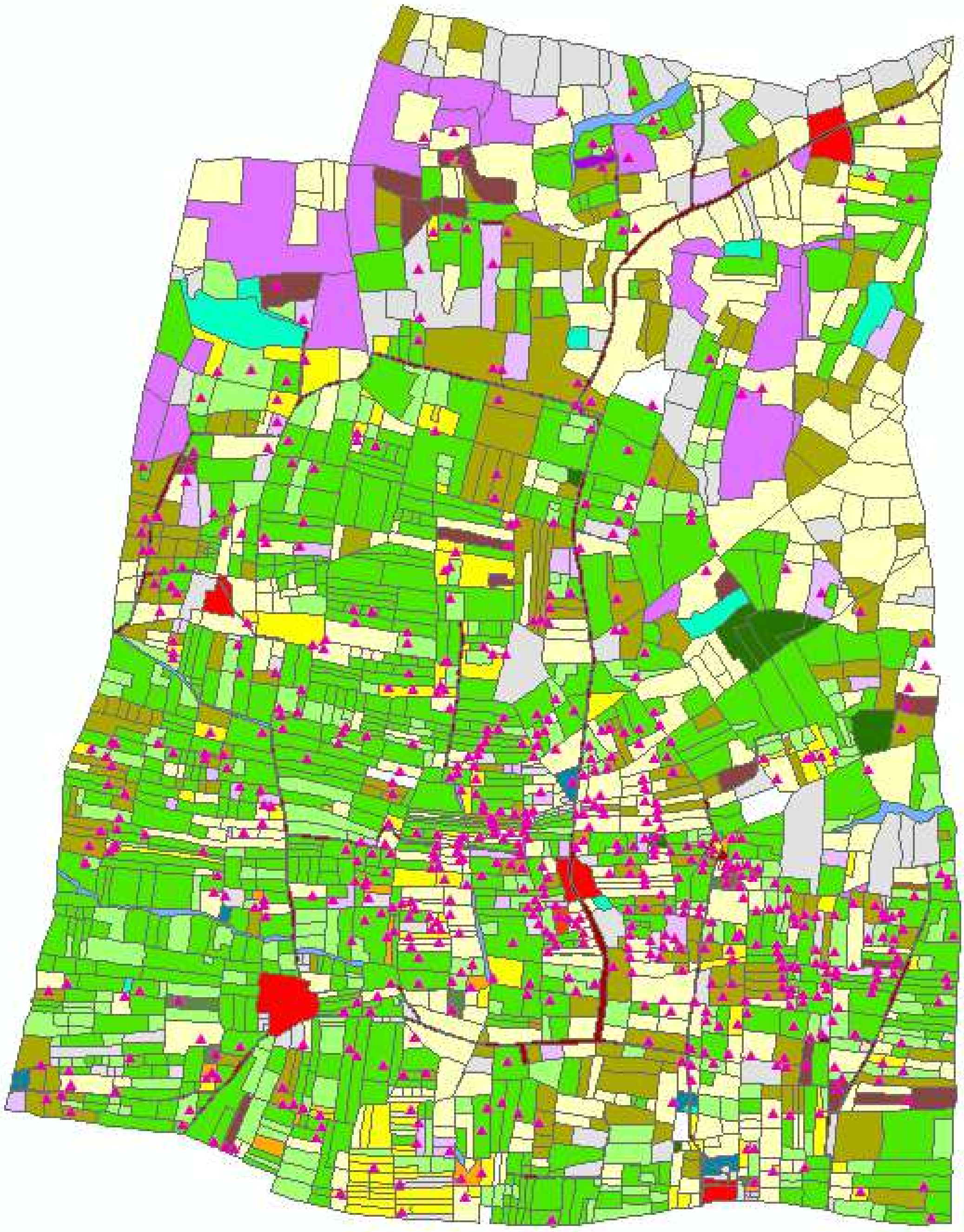


Figure 8. GPS locations traversed during November 2011

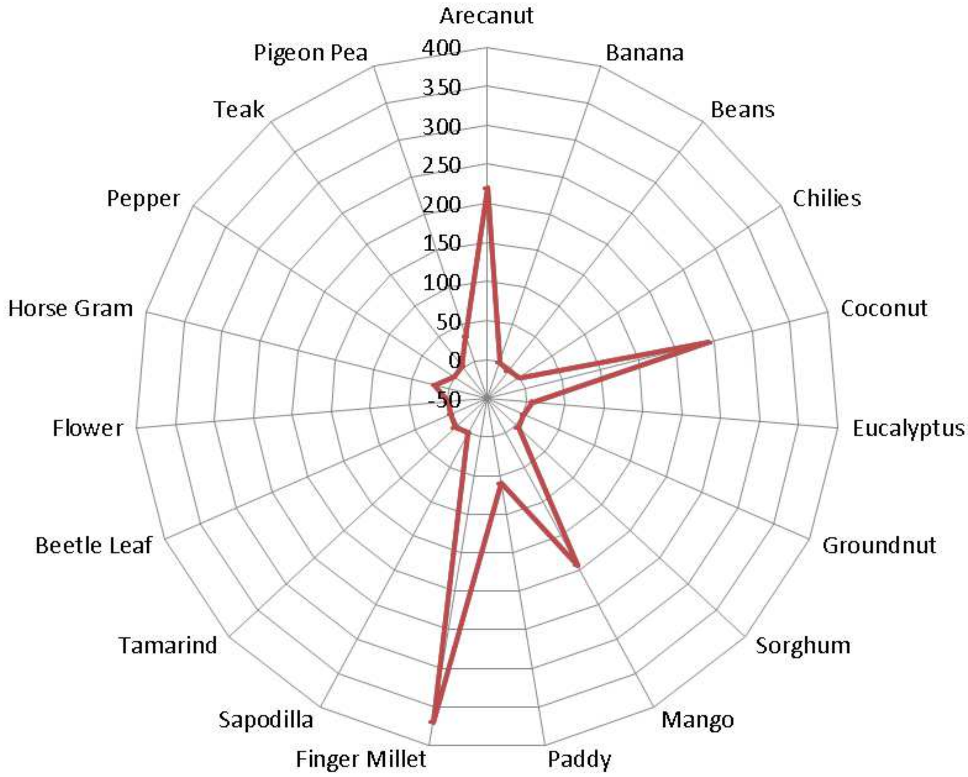


Figure 9. Difference in crop area obtained by conventional (c) and alternative methods (a) for the year 2011

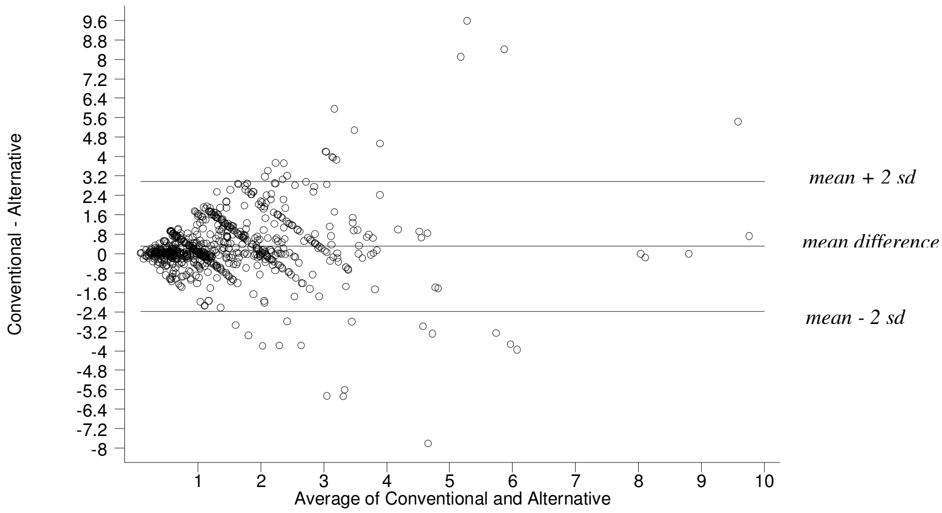


Figure 10. Difference in methods against their mean for total crop area (All crop)

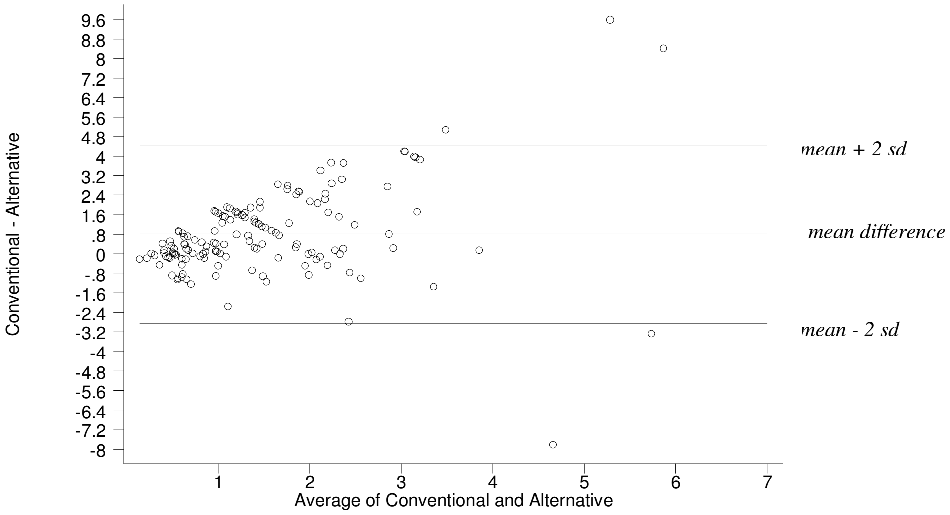


Figure 11. Difference in methods against their mean for arecanut crop area

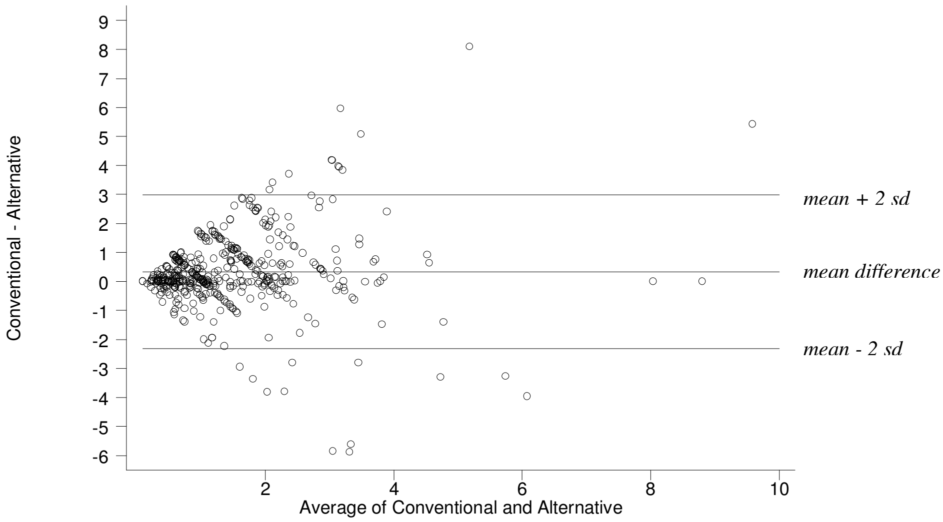


Figure 12. Difference in methods against their mean for coconut crop area

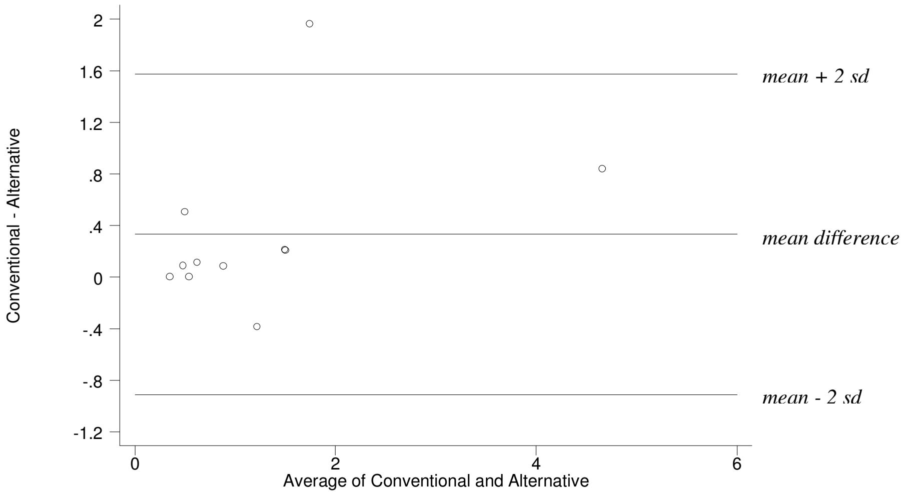


Figure 13. Difference in methods against their mean for sorghum crop area

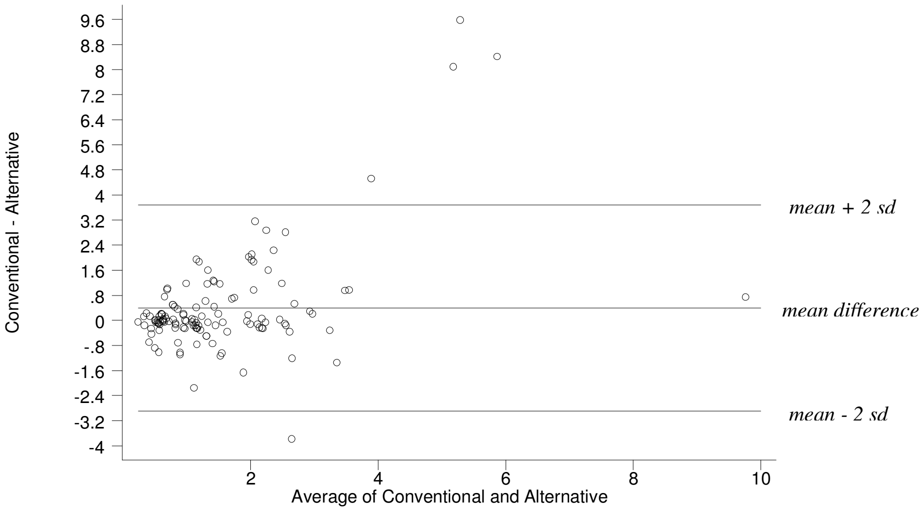


Figure 14. Difference in methods against their mean for mango crop area

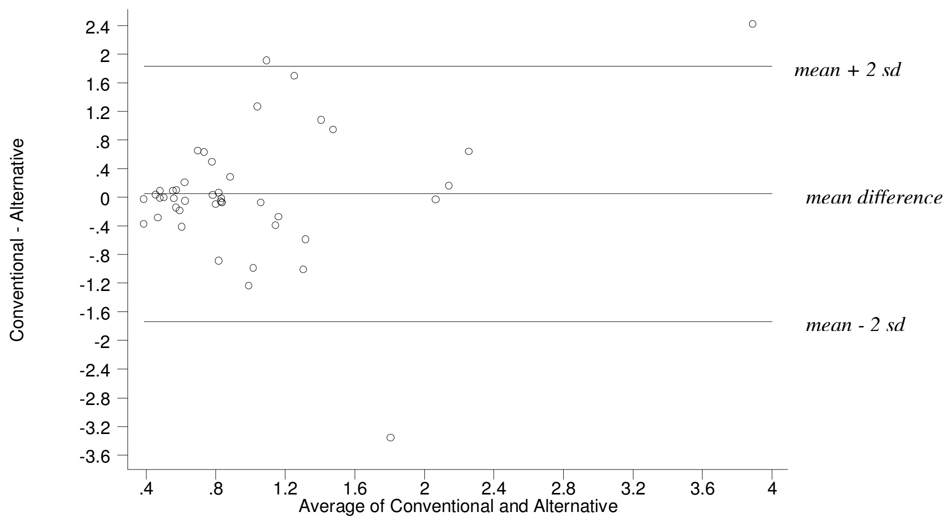


Figure 15. Difference in methods against their mean for paddy crop area

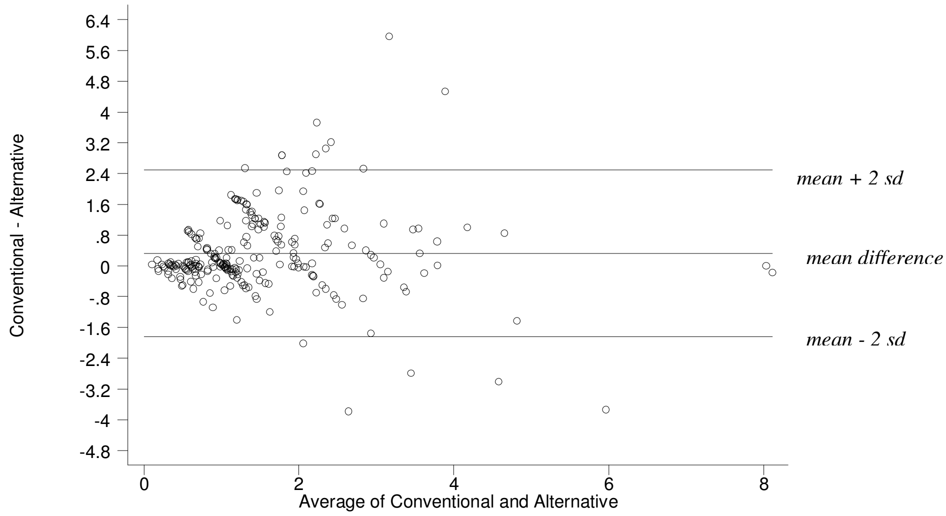


Figure 16. Difference in methods against their mean for finger millet crop area