Article

A Survey Bias Index Based on Unmanned Aerial Vehicle Imagery to Review the Accuracy of Rural Surveys

Xueyan Zhang

Key Laboratory of Land Surface Pattern and Simulation, Institute of Geographical Science and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China; xyzhang@igsnrr.ac.cn

Abstract: Field surveys and questionnaires are a cornerstone of rural socioeconomic research, providing invaluable firsthand data regarding on-the-ground situations. However, cost-effective and efficient methods for validating the accuracy of self-reported data in such questionnaires are lacking. Biased data are likely to lead to incorrect conclusions. In this study, we propose a new index, the survey bias index (SBI), for evaluating the degree of survey bias in field surveys. This index was obtained by comparing the data recorded in questionnaires with those from portable unmanned aerial vehicles (UAVs). In a case study, we employed SBI to reveal the degree of survey bias of questionnaires in field surveys on rural homesteads. The SBI of self-reported areas of rural homesteads reached 0.439, implying that 43.9% of data were significantly different from those collected using UAVs. A greater SBI was obtained in the pre-urban zone (0.515) than in the pure rural zone (0.258). These results indicate that homestead areas in the pre-urban zone have more incentive to expand than those in the pure rural zone. UAV remote sensing can strongly support research in the field of social economy, which reveals key information hidden in field surveys and questionnaires.

Keywords: survey bias; rural survey; unmanned aerial vehicle; data validation; China

1. Introduction

Field surveys are a cornerstone of rural socioeconomic research, providing invaluable firsthand data regarding on-the-ground situations. Questionnaires are a key component of field surveys that usually provide a low-cost method of obtaining information regarding a diverse range of factors. Questionnaire accuracy depends upon valid responses from interviewees. This can be ensured by careful questionnaire design as well as formal assessments of questionnaire repeatability and validity [1]. However, in practice, questionnaires are often used without thorough evaluation, which can result in unknown bias affecting results [2].

Survey bias is apparent not only for stigmatized or illicit behaviors [3,4] but also for minimally controversial topics, such as income [5], attending religious services, and voting [6]. Despite the importance of this concept, there have been few attempts to overcome this confidence problem. One solution to this problem is to introduce third-party data. Kar et al., compared data on self-reported consumption and purchases of LPG cylinders to assess survey biases regarding a clean fuel program in Karnataka, India; a large significant difference was observed between the self-reported mean annual consumption (3.53 cylinders) and annual consumption determined by purchase records (1.73 cylinders). Thus, only 15% of consumers accurately reported their consumption [7]. Tur-Sinai et al., compared self-reported dwelling prices and data from national samples of housing sale transactions using a census tract. They found that the self-reported estimates of dwelling values were, on average, 20% higher than the mean market prices of houses in the corresponding census tracts. The second solution is to compare results obtained using measuring devices [8]. To determine the accuracy of self-reported body mass index, self-reported and measured values were compared. Women tended to underestimate their weight, whereas men overestimated their weight [9]. Verbeij et al. compared the self-reported and the measured use of...
social media by adolescents and found that these subjects overestimated their time spent on social media [10]. To measure the confidence of the respondents, VR equipment with an electroencephalogram sensor and eye-tracking camera was used to assess social desirability bias from questionnaires. This pioneering strategy was used to incorporate both virtual reality and biosensor technology in surveys [11]. Based on these results, survey bias in questionnaires is quite common. The introduction of third-party data or that measured using devices has explained some survey bias. However, there is no uniform approach for representing the degree of bias.

Recently, China’s rural population has declined by approximately 13%, whereas the population in rural homestead areas has increased by approximately 4% [12]. The vacancy rate of rural homesteads in China has reached 10–15% [13], resulting in a large quantity of unused rural land [14]. Consequently, rural homesteads have attracted the attention of rural social researchers [15]. Questionnaires administered via personal interviews can be applied in rural homesteads for exploring the informal withdrawal homestead issue [16]. Because homesteaders often have strong incentives to misstate data (e.g., to avoid revealing illegal or informal housing arrangements or to receive increased government subsidies) [17–19], questionnaires can lead to highly biased results depending on the status of the respondents. To overcome these limitations, new methods for rural homestead surveys should be developed to minimize survey bias. This research area presents an ideal opportunity for testing a new method for identifying survey bias in reported data [20].

This study presents a new index method implementing new technologies, such as an unmanned aerial vehicle (UAV), specifically for evaluating survey bias encountered during rural surveys. Survey bias can be revealed by comparing data collected on homesteads from questionnaires; UAVs can be used to verify the reliability of rural homestead survey data. This research attempts to fill two gaps by (1) introducing UAVs as measuring devices for verifying the survey bias of self-report homestead areas in rural surveys; and (2) developing a new index method for identifying the degree of survey bias and enabling comparison of the bias results obtained from different areas. Therefore, this research provides academic and practical insights into how the survey bias index (SBI) based on UAVs can reveal attitude responses in rural surveys and how more information can be mined for rural management.

To present the application of this new index method based on UAV, we selected Jianfeng Village in Anhui province in China as the study case. This paper is divided into five sections. The next section explains the case selection, and it elaborates on the methodological aspects of this paper. Section 3 presents the results from the application of the SBI. Section 4 discusses the results, and Section 5 presents the main conclusions.

2. Methods

2.1. Profile of the Study Area

Anhui Province is the pilot location of China’s systemic reform of collective rural property rights. Qimen County is globally renowned for its black tea. Jianfeng Village, located along the eastern edge of Qimen County in the town of Qishan, is a typical area of rural–urban transfer (Figure 1). We selected two zones in Jianfeng Village to test our questionnaire and UAV survey: the pre-urban area (Zone A) and a strictly rural area (Zone B).
2.2. Questionnaire and Survey Design

Field survey interviews with farmers residing in 31 different rural homesteads were conducted over a 3-day field survey. Each interview lasted for approximately 30 min. The interviews comprised questions about the farmers’ homesteads, including the area of different functional parts (e.g., main building, ancillary building, and yard), as well as the year of construction, number of floors, and number of residents (Table 1). Interview data from each farmer were coded as N01–N31. Homesteads N01–N12 are in Zone A, whereas N13–N31 are in Zone B. The investigator explained that the purpose of the survey was to evaluate the accuracy of the homestead area, and the survey data would be compared with UAV remote sensing data. Because the study area had many rural residential reconstructions and expansions, and differences exist between the homestead area and the registered area, we chose to record the homestead area data as self-reported in this study.

Table 1. Questionnaire on rural homestead.

<table>
<thead>
<tr>
<th>Number</th>
<th>Question</th>
<th>Options or Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>Your gender</td>
<td>1 = male; 0 = female</td>
</tr>
<tr>
<td>Q2</td>
<td>Your age</td>
<td>Years</td>
</tr>
<tr>
<td>Q3</td>
<td>How many years of school education have you received?</td>
<td>Years in school</td>
</tr>
<tr>
<td>Q4</td>
<td>Your employment</td>
<td>1 = agriculture; 0 = non-agriculture</td>
</tr>
<tr>
<td>Q5</td>
<td>What is the area of the homestead</td>
<td>Area of homestead, m²</td>
</tr>
<tr>
<td>Q6</td>
<td>How many people currently live in the household?</td>
<td>Number of individuals</td>
</tr>
<tr>
<td>Q7</td>
<td>How many years has the residence been built?</td>
<td>Years</td>
</tr>
</tbody>
</table>
2.3. UAV Survey

Most studies of UAV technology have focused on the accuracy of data collection, whereas few studies have focused on the associated attitudes of owners of rural homesteads. For instance, Yang et al., used Dajiang UAVs to measure the building density and floor area ratio of rural settlements [21]. UAV can simultaneously acquire real-time high-resolution images, offset the fixed acquisition cycle of optical sensor satellite remote sensing, and respond to dynamic weather conditions (e.g., cloud cover) [22].

A DJI MAVIC Pro UAV (Da-Jiang Innovations, Guangdong, China) was used to estimate the area of each homestead. This UAV was a quadcopter featuring four electric motors, and it was equipped with a CMOS camera with a focal length of 28 mm and 12.35 million effective pixels in a 1-inch CMOS. The UAV’s maximum speed was 18 m/s, and the maximum flight duration was 21 min. UAV data were acquired on 16 August 2019. A dry, non-windy day was selected to avoid distortions caused by UAV camera undulations. Autonomous flight planning was conducted across the study area. The flight path was oriented in the north–south direction at a speed of 11 m/s in two flights of approximately 12 min each. The camera was set to a F-stop of 2.2, shutter speed of 1/2000 s, and ISO-values that adapted automatically to available light conditions, between 100 and 1600 ISO. During UAV image acquisition, a camera angle of −90° was used, and safe mode was engaged. The sensor produced three 20-MP images in the red, green, and blue (RGB) wavelengths. During the flight, the camera was automatically released every 7 m, while the device’s position was simultaneously recorded by an internal GPS/GLONASS dual-mode satellite positioning system. In Zone A, the flight longitude ranged from 117°42′57.60″ E to 117°43′12.00″ E, and the flight latitude ranged from 29°51′07.20″ N to 29°51′10.80″ N. Images were obtained from a height of 109 m with a 70% side overlap, a 90% forward overlap, and an optical GSD of 3.4 cm. The flight area covered was 66,394.21 m². In Zone B, the flight longitude ranged from 117°42′57.60″ E to 117°43′12.00″ E, and the flight latitude ranged from 29°51′07.20″ N to 29°51′10.80″ N. Images were obtained from a height of 100.4 m with a 70% side overlap, a 90% forward overlap, and an optical GSD of 3.1 cm. The flight area covered was 61,196.23 m². During the survey, 130 RGB images were obtained.

2.4. Three-Dimensional Modeling for Visual Interpretation

The structure from the motion method was used to build a 3D model to enhance the visual interpretation of the collected data. The workflow of the structure from the motion method consists of two major processes: aligning images and building geometry. The 3D point cloud was generated using Agisoft Photoscan Professional Edition software (Agisoft LLC, St. Petersburg, Russia) [23]. First, the camera position for each image and four common points in the images were located and matched to enable the identification of calibration parameters for image comparison. The point cloud was built based on the estimated camera positions and images [24]. A digital surface model and an orthomosaic were also generated [25]. The 3D point cloud digital surface model and the orthophoto are shown in Figure 2a. The orthophoto was used for visual interpretation, with 1-pixel precision in ArcGIS. The 3D tiled model assisted in identifying different functional areas of each rural homestead, such as the main building, ancillary building, and yard (Figure 2b).
Figure 2. Three-dimensional model to enhance visual interpretation: (a) three-dimensional tiled model of Zone B and (b) 3D model helps to identify different functional areas of rural homestead from orthographic and different perspectives.

2.5. Survey Bias Index

Values obtained from the UAV were set as the baseline. Survey bias was then estimated using the degree of shift from the baseline. Therefore, in the absence of survey bias, the homestead areas reported by the respondents should closely match the information extrapolated from UAV data. Any discrepancy between the datasets indicated some degree of survey bias in the household survey data. To minimize error from respondents’ inaccurate memory, the measurement at the main building area was selected for comparison between the survey data and the UAV data (as respondents were likely to have the greatest familiarity with the dimensions of this facility).

To evaluate survey bias during surveys on rural homesteads and to analyze the interspecific differences in the two survey methods, the index of survey bias for rural surveys was calculated using the least squares method:

\[
SBI = \frac{\sum_{i=1}^{n} (V_{Gap,i} - V_{Gap}) (V_{UAV,i} - V_{UAV})}{\sum_{i=1}^{n} (V_{Gap,i} - V_{Gap})^2} \tag{1}
\]

where \(SBI\) is the survey bias index, \(V_{UAV}\) is the value associated with the rural homestead obtained by UAV (m\(^2\)), \(V_{Gap}\) is the difference between homestead areas obtained from self-reported questionnaires and UAV (m\(^2\)), \(n\) is the number of rural homesteads surveyed using the two methods, and \(i\) is the \(i\)-th zone surveyed.
An independent samples t-test was used to analyze the mean value comparison. All analyses were performed in SPSS version 19 software (SPSS, Inc., Chicago, IL, USA), and the results were considered significant when $p \leq 0.05$.

3. Results

3.1. Descriptive Statistics

Most responses to the survey were from 31 households (Table 2). Male respondents accounted for 58%, and female respondents accounted for 42%. Most residents of the surveyed rural households graduated from primary school or more (79.31%), and only 31.03% of the surveyed sample insisted on engaging in agricultural production. The average area of the homestead was 90.48 m$^2$, and many households had more than three family members (72.41%).

Table 2. Summary statistics of variables used in analysis.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Average</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender (1 = male; 0 = female)</td>
<td>0</td>
<td>1</td>
<td>0.58</td>
<td>0.49</td>
</tr>
<tr>
<td>Age (years)</td>
<td>32</td>
<td>82</td>
<td>58.93</td>
<td>13.01</td>
</tr>
<tr>
<td>Education level (years in school)</td>
<td>0</td>
<td>12</td>
<td>6.31</td>
<td>3.96</td>
</tr>
<tr>
<td>Employment (1 = agriculture; 0 = non-agriculture)</td>
<td>0</td>
<td>1</td>
<td>0.31</td>
<td>0.46</td>
</tr>
<tr>
<td>Area of homestead (m$^2$)</td>
<td>50</td>
<td>170</td>
<td>93.20</td>
<td>29.13</td>
</tr>
<tr>
<td>Number of individuals in the household</td>
<td>1</td>
<td>10</td>
<td>4.10</td>
<td>2.27</td>
</tr>
<tr>
<td>Years for residence been built (years)</td>
<td>1</td>
<td>64</td>
<td>29</td>
<td>13.97</td>
</tr>
</tbody>
</table>

3.2. Areas of Homesteads Identified by Questionnaire and UAV

Twelve rural homesteads were surveyed using a questionnaire in Zone A, a typical pre-urban zone (Figure 3a). The areas of rural homesteads in Zone A range from 65.12 to 243.11 m$^2$, with a mean of $138.57 \pm 58.69$ m$^2$ (Figure 3a). Existing buildings in Zone A form the original neighborhood. Most buildings are 2–4 stories and self-built by farmers for rental purposes. Figure 3b shows all areas of the homesteads in Zone A as a visual interpretation based on UAV data. Nineteen rural homesteads were surveyed using a questionnaire in Zone B (Figure 3c). Their areas range from 50.16 to 163.88 m$^2$, with a mean of $105.98 \pm 32.11$ m$^2$ (Figure 3c). The buildings in Area B are distributed along the riverbank. There are 1–2 story buildings occupied by farmers for personal use. Whole homesteads in Zone B were interpreted based on orthographic and 3D UAV data (Figure 3d).

3.3. SBI in the Rural Homesteads Survey

As shown in Figure 4, the mean of homestead areas obtained by UAV was larger than that determined from the questionnaire ($N = 31, p < 0.01$). According to the questionnaire data, the area of homesteads ranged from 50.00 to 170.00 m$^2$, with a mean value of $93.20 \pm 29.13$ m$^2$. In contrast, the area of rural homesteads surveyed by UAV ranged from 50.16 to 243.11 m$^2$, with a mean value of $118.60 \pm 46.28$ m$^2$. The difference in mean areas of rural homesteads obtained using the UAV and questionnaire was up to 25.40 m$^2$ ($N = 31, p < 0.01$), which could not be attributed to random error.
Figure 3. Area of rural homesteads surveyed by questionnaire and unmanned aerial vehicle (UAV) in Zones A and B: (a) distribution of rural homesteads surveyed through a questionnaire in Zone A; (b) rural homesteads interpreted by UAV in Zone A; (c) distribution of rural homesteads surveyed through a questionnaire in Zone B; and (d) rural homesteads interpreted by UAV in Zone B. The orange dotted line is the scope of the survey. The yellow dotted line is the questionnaire survey path.
According to Formula 1, the overall SBI of the surveyed homesteads was 0.439, and the total areas of the homesteads were 3676.60 m² ($V_{UAV}$) and 2889.41 m² ($V_{questionnaire}$). This represented a slope between $V_{Gap}$ and $V_{UAV}$, reflecting the ratio of the difference between areas obtained using the UAV and the questionnaire to actual areas of the homesteads. The areas of homesteads obtained in Zone A using the UAV and the questionnaire were 1652.12 and 1237.29 m², respectively. The areas of homesteads obtained in Zone B using the UAV and the questionnaire were 2013.75 and 1662.85 m², respectively. Formula 1 gave an SBI of 0.515 in Zone A and 0.258 in Zone B (Table 3). Thus, the SBI was greater in the pre-urban zone than in the pure rural zone.

As shown in Figure 5, a significant positive correlation was observed between the difference in the areas measured by the UAV and the questionnaire survey for the property size (N = 31, $p < 0.001$). As the areas of the homesteads increased, the difference in the results between the questionnaire and the UAV also increased. This result reveals that respondents with larger homesteads show greater data bias than others.

**Table 3.** Survey bias index (SBI) of rural homestead survey in the questionnaire and UAV. $V_{UAV}$ is the area of a rural homestead obtained by the UAV (m²), $V_{questionnaire}$ is the area of a rural homestead obtained by the self-reported questionnaire (m²), and $n$ is the number of rural homesteads surveyed (12 and 19 in Zones A and B, respectively).

<table>
<thead>
<tr>
<th>Survey Areas</th>
<th>$\sum_{i} V_{UAV,i}$ (m²)</th>
<th>$\sum_{i} V_{questionnaire,i}$ (m²)</th>
<th>$\sum_{i} V_{Gap,i}$ (m²)</th>
<th>SBI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zone A</td>
<td>1652.12</td>
<td>1237.29</td>
<td>425.56</td>
<td>0.515</td>
</tr>
<tr>
<td>Zone B</td>
<td>2013.75</td>
<td>1662.85</td>
<td>361.63</td>
<td>0.258</td>
</tr>
<tr>
<td>Study area</td>
<td>3676.60</td>
<td>2889.41</td>
<td>787.19</td>
<td>0.439</td>
</tr>
</tbody>
</table>
Figure 5. Positive correlation between the difference in the areas measured from the survey regarding the property size. Gray lines are regression lines with a 95% confidence interval.

4. Discussion

Field surveys and questionnaires are a cornerstone of rural social and economic research, and they provide invaluable firsthand data regarding on-the-ground situations. However, cost-effective and efficient methods for validating the accuracy of self-reported data from such questionnaires are lacking. In this study, we propose the SBI, which evaluates the degree of survey bias in field surveys on rural homesteads by comparing data recorded from questionnaires and portable UAV. The calculated SBI revealed that the degree of survey bias of questionnaires from field surveys on rural homesteads reached 0.439, reflecting that the ratio of the difference between areas obtained using the UAV and questionnaire to actual areas of the homesteads reached 43.9%. Survey bias has three possible sources: (1) Different measurement methods lead to different measurement areas; (2) The measurement methods are the same, but accuracy errors exist; and/or (3) Inconsistent area due to house reconstruction and expansion [26].

The SBI obtained in the pre-urban zone (0.515) was higher than that in the pure rural zone (0.258). The homestead areas in the pre-urban zone showed more survey bias than those in the pure rural zone (Table 3). In addition to the three reasons mentioned above, the homesteads in the pre-urban zone have greater possibilities and a higher frequency of reconstruction and building extension for rental and other businesses [18]. Therefore, the homestead areas in the pre-urban zone showed a greater SBI than those in the pure rural regions—where there is a lack of such motivation to reconstruct—and the degree of survey bias was significantly lower. As shown in Figure 5, the larger homesteads showed greater data bias than others. One possible explanation is that the rebuilt houses may expand the area, resulting in a great survey bias in the investigation of a large area of the homestead.
Another rural homestead survey also reported that the evolution and the upgrading of the structure and the function of homesteads are influenced by factors such as expanding production, improved living conditions, promotion of regional economic development, and mutual comparison between neighbors [15].

This method can be widely applied to many other survey indicators that can be obtained through remote sensing. UAV is a real-time, in situ survey tool that can interpret many land features, including crop area, crop yield, quantity of large agricultural machinery, types of cash crops, soil fertility, and forest resources, all of which are common targets of rural research. Furthermore, compared with RGB cameras, UAVs equipped with multispectral sensors can investigate the rural environment, such as the water quality of rural rivers. Therefore, SBI is an efficient method for quickly judging the data quality of rural surveys and research, which should be confirmed in further expanded research.

The SBI method based on UAV has some limitations. First, it can only be used to verify specific objects in rural surveys, rather than abstract objects, such as income, employment, and education. It cannot be used for objects invisible to UAVs, such as indoor assets. However, in this case, this part of the data can be used to verify the reliability of the questionnaire data. Second, the duration of a UAV survey is limited. Periodic UAV data acquisition should be used to construct a long-timescale continuous rural panel data set [27]. Third, the scope and the scale of UAV surveys are limited. When surveys are conducted in large villages, small drones may not be able to obtain data effectively.

5. Conclusions

This study proposed an SBI for evaluating the accuracy of self-reported rural surveys using questionnaires based on a comparison of data recorded from questionnaires and a portable UAV. The calculated SBI revealed that the degree of survey bias of questionnaires in field surveys on rural homesteads reached 0.439, reflecting that the ratio of the difference between areas obtained using the UAV and questionnaire to actual areas of the homesteads reached 43.9%. A greater SBI was obtained in the pre-urban zone (0.515) than in the pure rural zone (0.258). Our results suggest that homestead areas in the pre-urban zone have more incentive to expand than those in the pure rural zone. The SBI based on UAV remote sensing can be used as an efficient tool for validating the quality of data and the extent of survey bias. Our case study shows that UAV remote sensing can strongly support research in a rural survey, revealing key information hidden in field surveys and questionnaires.

Funding: National Key R&D Program of China (No. 2021YJ0093, No.2018YFC1508805, and No. 2019YFA0607403), Central Public Interest Scientific Institution Basal Research Fund (No. Y2020PT05 and No. BSRF201901), and the Strategic Priority Research Program of Chinese Academy of Sciences (XDA20010302).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on requested from the author.

Conflicts of Interest: The author declares no conflict of interest.

Declaration of Self-Responsibility for Ethical Issues: The author declares that she will take responsibility for any ethical issues in this paper.

References


17. Ward, P.M.; Peters, P.A. Self-help housing and informal homesteading in peri-urban America: Settlement identification using digital imagery and GIS. Habitat Int. 2007, 31, 201–218. [CrossRef]


