

Landscape diagnostic survey data of wheat production practices and yield in eastern India

Anurag Ajay^{1*}, Peter Craufurd², Sachin Sharma¹, Harshit Ranjan¹, Gokul Poudel², Ram Kanwar Malik³, Balwinder Singh³, Ashok Kumar Singh⁴, Arindam Samaddar⁵, Ashok Rai⁶, Alwin Keil⁷ and Andrew McDonald⁸

¹ International Maize and Wheat Improvement Center (CIMMYT), Patna, Bihar, India

² International Maize and Wheat Improvement Center (CIMMYT), Lalitpur, Kathmandu, Nepal

³ International Maize and Wheat Improvement Center (CIMMYT), Pusa, New Delhi, India

⁴ Indian Council of Agricultural Research (ICAR), Division of Agricultural Extension, New Delhi, India

⁵ International Rice research Institute (IRRI), Pusa, New Delhi, India

⁶ Danish Refugee Council, Bahan Township, Yangon, Myanmar

⁷ UNIQUE forestry and land use GmbH, Schnewlinstr. 10, 79098 Freiburg, Germany

⁸ Cornell University, College of Agriculture & Life Sciences, Ithaca, NY 14850 USA

*email: a.ajay@cgiar.org

Abstract: Approximately 7,600 wheat plots were surveyed and geo-tagged in the 2017-18 winter or *rabi* season in Bihar and eastern Uttar Pradesh (UP) in India to capture farmers' wheat production practices at the landscape level. A two-stage cluster sampling method, based on Census data and electoral rolls, was used to identify 210 wheat farmers in each of 40 districts. The survey, implemented in Open Data Kit (ODK), recorded 226 variables covering major crop production factors such as previous crop, residue management, crop establishment method, variety and seed sources, nutrient management, irrigation management, weed flora and their management, harvesting method and farmer reported yield. Crop cuts were also made in 10% of fields. Data were very carefully checked with enumerators. These data should be very useful for technology targeting, yield prediction and other spatial analyses.

Keywords: agronomy, landscape diagnostic survey, production practices survey, wheat and yields.

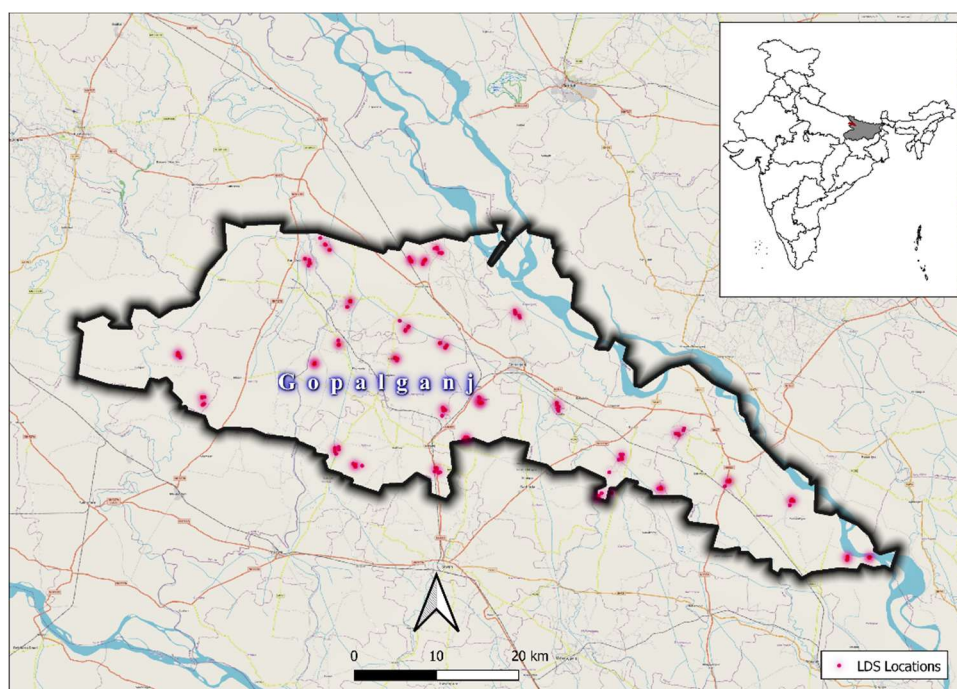
1 INTRODUCTION: Crop yields are known to vary widely in space and time, and the so called yield-gap (defined here as the difference in yield between the best and worst 10% of farmers) can be substantial (e.g. Global Yield gap Atlas <http://www.yieldgap.org/web/guest/home>). Closing this yield gap through good agronomic practices or best management practices is the aim of most extension programs. But what are the best management practices being used by farmers? How do they vary spatially (and temporally)? Are they predictable? The Landscape Diagnostic Survey (LDS), which is being implemented in wheat and rice systems in eastern India, is designed to capture farmers' current practices for cultivating wheat and rice at the scale of large landscapes, and to use these data, and other spatial data, to understand and predict yield and the key drivers of production.

The Cereal System Initiative for South Asia (CSISA: <https://csisa.org/>), in collaboration with the Indian Council of Agricultural Research (ICAR <https://icar.org.in/>) and State Agricultural Universities (SAU), surveyed wheat farmers in Indian states of Bihar and eastern UP to capture their current production practices. ICAR and SAUs between them have an extensive network in the field through the *Krishi Vigyan Kendra* - KVK system (<https://kvk.icar.gov.in/>). All partners jointly developed the survey questionnaire and methodology. The survey covered all aspects of production, including agronomic, social, economic and market variables, assuming that yield might also depend on other factors besides agronomic management. The sampling methodology was devised to ensure a representative sample keeping in mind the cost and time involved. Special emphasis was given on randomized selection of farmers and their spatial distribution. Farmers were interviewed individually and their production practices on the largest wheat plot were recorded for the winter or *rabi* season 2017/2018. Physical crop-cuts were also planned for a sub-set of samples to check the deviation between survey reported and crop-cut yields. The survey aimed to generate data based evidence around current crop production practices that can be wisely utilized by national and state level policy makers for enhancing crop productivity in the region.

2. FIELD SURVEY AND DATA COLLECTION

2.1 Sampling method: Two-stage cluster sampling was applied to ensure a balance between available resources and desired accuracy (Sedgwick, 2014). A District was considered as a survey unit and villages as clusters within a District. In the first stage, villages were selected through probability proportionate to size (PPS) method as villages vary in size. Larger villages were assigned higher probability of selection than smaller villages (Skinner, 2020). In the second stage, the same number of households (HHs) were selected randomly in each sampled village so that each unit sampled had equal chance of getting selected. Village selection was performed using data from '2011 Census of India'. All villages within a district were enlisted along with their sizes (number of HHs). Villages listed under 'urban' category, having more than 5000 HHs (extremely big) and having less than 50 HHs (extremely small) were removed. The remaining villages formed the sampling frame for village selection. PPS was applied on this frame to draw 30 villages randomly. Farmer selection relied on the 'list of voters' fetched from State's Election Commission website. The village list provided the names of all residents along with unique house numbers. These house numbers were used to construct the sampling frame for HH selection. From each sampled village, seven HHs were selected using simple random sampling. Accordingly, 210 HHs were interviewed in each District. An example of sample distribution in Gopalganj District of Bihar is portrayed (Figure 1).

Figure 1: Dots showing geo-location of surveyed farmers' plot in Gopalganj District of Bihar



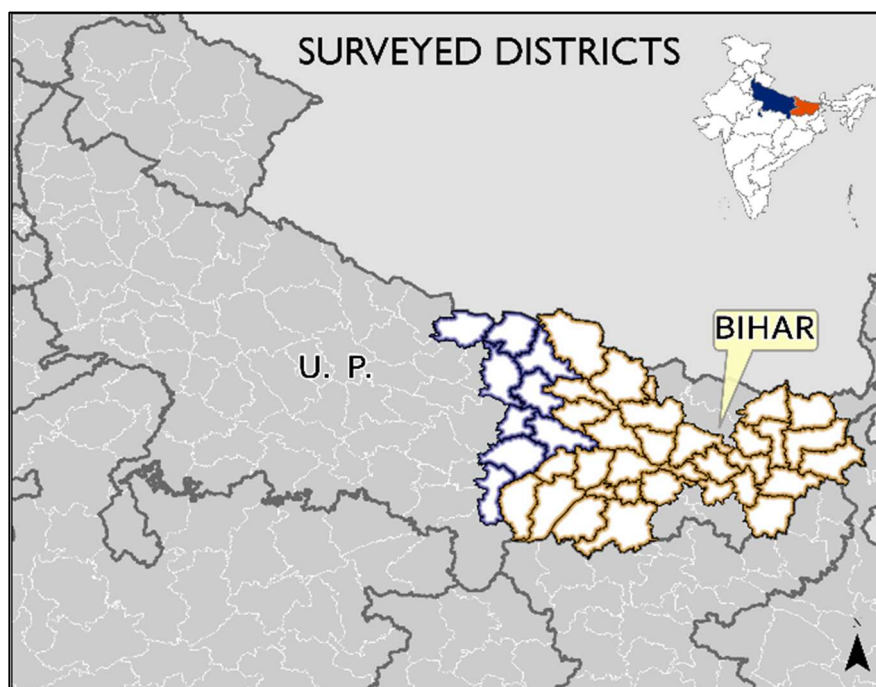
2.2 Digital survey instrument and ODK tool: The survey was deployed electronically using ODK. This enabled real-time progress monitoring, automation in data compilation and error minimization during interviews. The questionnaire was programmed in an offline version (.xlsx version) of ODK Build. The survey instrument had been refined over a number of cycles such that there were no open-ended questions and minimum-maximum ranges were applied to reduce errors in entering values. Enumerators used ODK Collect, an Android application (App) to capture interview responses. Raw data sent by enumerators was stored on ODK Aggregate, an open source Java App which also hosted the blank questionnaire in XForm version. Enumerators downloaded blank questionnaire on their Android devices, completed interviews and sent back filled-in questionnaires at Aggregate (<https://docs.getodk.org>). The survey instrument and the ODK version (XML) are included in the downloadable files.

2.3 Survey deployment: The survey was deployed through staff of the *Krishi Vigyan Kendras* (KVK), a Government agricultural extension centre in each District. Concerned staffs of all 40 centres attended 'orientation on sampling method, survey questions and training on application of ODK in four separate batches, comprising one day of classroom training followed by mock interviews of farmers using ODK

Collect App on next day. Participants were provided with the list of sampled villages and respondents of their respective Districts. During survey deployment, they received technical and logistical support from the project.

2.4 Coverage: The survey covered 40 Districts and 7648 wheat farmers in Bihar and the eastern UP. All these districts together form a large area in eastern Indo-Gangetic plain of India (Figure 2) where the rice-wheat cropping system prevails. There were 31 Districts with 5793 farmers from Bihar, and nine districts with 1855 farmers from UP. The survey was conducted on the selected farmer's largest wheat plot.

Figure 1: Part of Indian map illustrating positions of surveyed Districts in Bihar and eastern UP



2.5 Data flow: Figure 3 shows the interlinkage among survey steps. The digital survey form (questionnaire) was designed on the .xlsx version of the ODK Build (1). Blank form was uploaded on the ODK server (2). Mobile devices linked with this server pulled blank forms for use (3). Selected farmers were interviewed (4) and completed forms were sent to the server (5). Raw data aggregated at the server was imported as .csv file (6). Data was curated by carefully screening and validating entries with the enumerators (7). The curated and cleaned file was analysed with open access software R to identify key yield attributing factors (8).

2.6 Data repository and format: The data is available from the CIMMYT CSISA Dataverse (<https://data.cimmyt.org/dataverse/csisadvn>). Data is available in an .xls file with metadata and variables, links to documents with the sampling method and survey instrument, and also the R script to read the data.

3. DATA SUMMARY: A summary of a few key agronomic variables are given in Table 1, although the survey has captured many other ecological, social, economic and market related parameters. The random sampling approach enabled us spread data across different land typologies as conceptualized by farmers (Figure 4). Seventy-one percent of data points were from medium land types, defined as lands that neither dry-up quickly nor face water logging situation after rain. Survey covered approximately 1100 villages and highlighted that 95% of the wheat plots are planted through broadcasting method. Rest 5% of the survey plots were line sown after tillage and under zero tillage in almost equal proportion. Planting time of wheat is an important variable whose influence on yield is well established (Malik et al., 2007). Wheat planting time in this part of India generally starts in the month of November and finishes by end of December. The survey captured this planting pattern;

Figure 3: Schematic diagram showing survey data flow

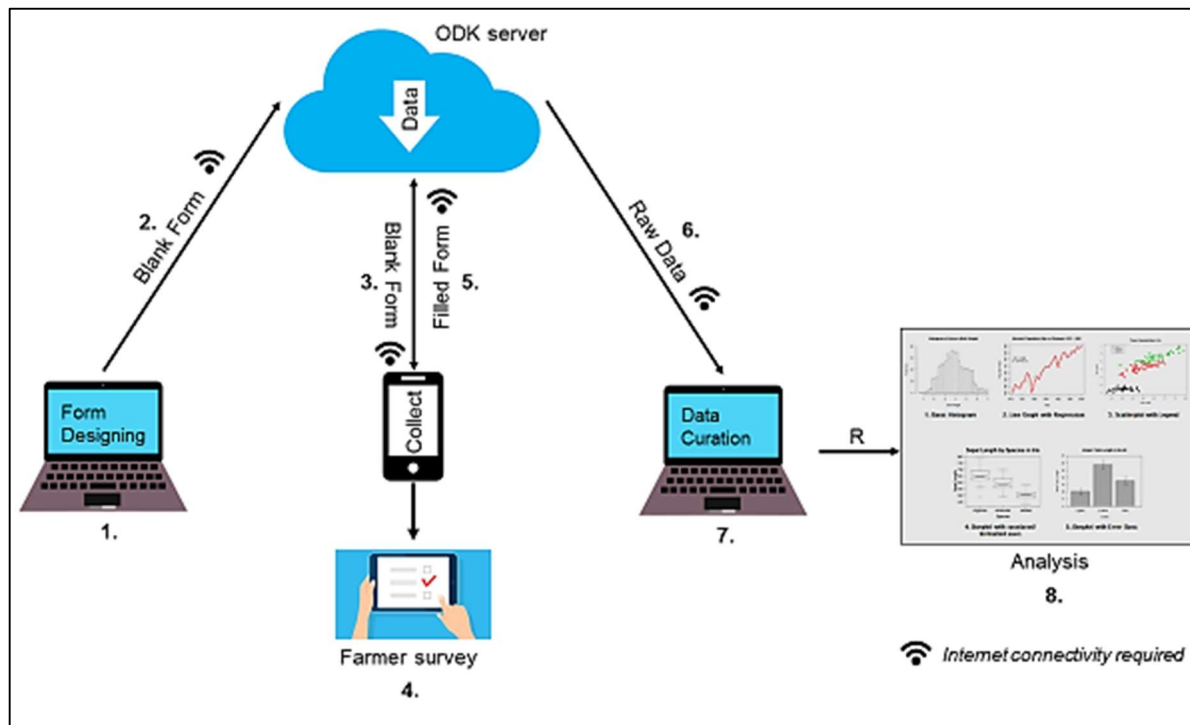
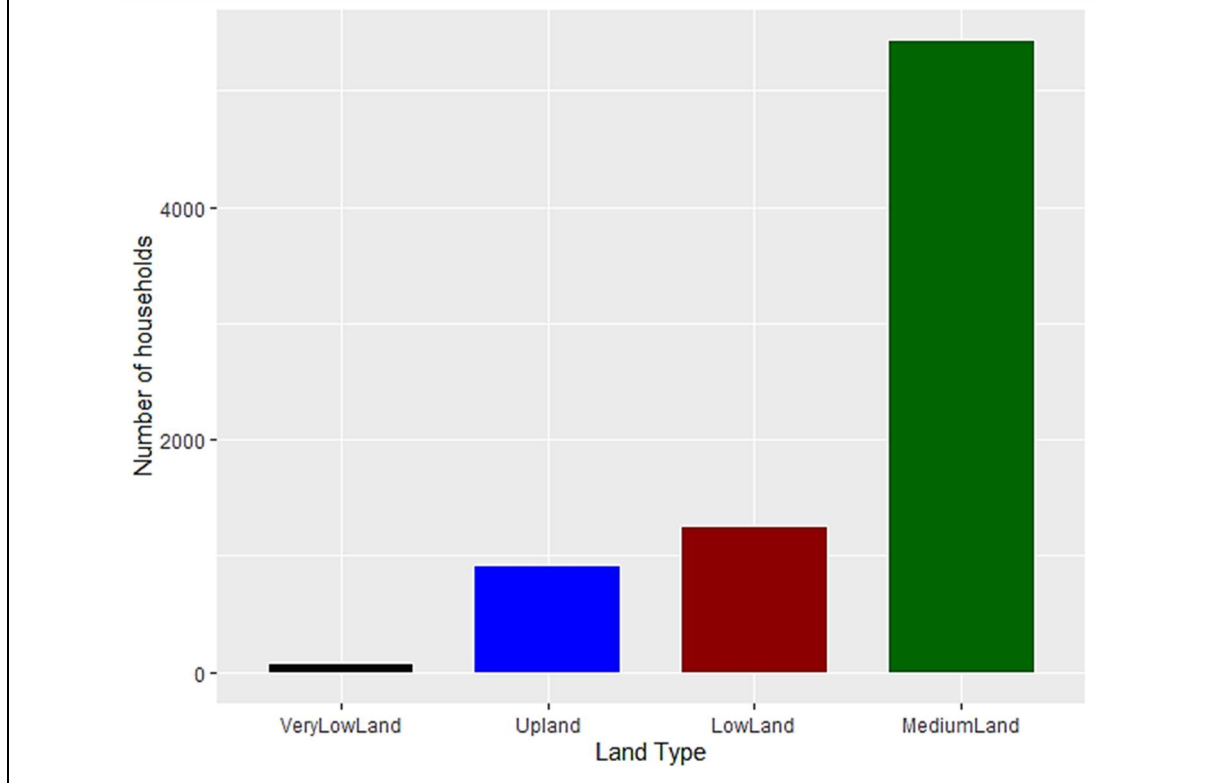


Figure 4: Land type of the largest plot of surveyed farmers



planting date ranged from 25 October to 26 January with a peak (28%) happening in the last week of November (Figure 5). Late sowing of wheat results in yield penalty (Singh et al., 2019). The survey categorically recorded reasons for delayed wheat planting wherever farmers had previously answered planting time after November. Farmers reported using 66 different wheat varieties but interestingly,

Table 1: Summary description of few major agronomic variables recorded in the survey

VARIABLES		Bihar N=5793	Uttar Pradesh N=1855
Wheat variety type planted	<i>Improved</i>	5539 (95.6%)	1688 (90.9%)
	<i>Local</i>	236 (4.1%)	153 (8.3%)
	<i>Unknown</i>	18 (0.3%)	14 (0.8%)
Crop establishment method	<i>Broadcasting</i>	5468 (94.4%)	1775 (95.7%)
	<i>Line sowing (after tillage)</i>	187 (3.2%)	35 (1.9%)
	<i>Zero tillage</i>	138 (2.4%)	45 (2.4%)
Mineral fertilizers applied	<i>No</i>	19 (0.3%)	3 (0.2%)
	<i>Yes</i>	5774 (99.7%)	1852 (99.8%)
Irrigation available	<i>No</i>	26 (0.4%)	1 (0.05%)
	<i>Yes</i>	5767 (99.6%)	1854 (99.9%)
Number of times irrigated ¹		2.26 (0.76)	2.38 (0.78)
Drought severity	<i>High</i>	173 (3.0%)	22 (1.2%)
	<i>Low</i>	1946 (33.6%)	312 (16.8%)
	<i>Medium</i>	1454 (25.1%)	714 (38.5%)
	<i>None</i>	2220 (38.3%)	807 (43.5%)
Weed severity	<i>High</i>	563 (9.7%)	298 (16.1%)
	<i>Low</i>	1679 (29.0%)	265 (14.3%)
	<i>Medium</i>	3245 (56.0%)	1055 (56.9%)
	<i>None</i>	306 (5.3%)	237 (12.7%)
Harvesting method	<i>Combine</i>	380 (6.5%)	821 (44.2%)
	<i>Manual</i>	5303 (91.5%)	1028 (55.4%)
	<i>Reaper</i>	110 (2.0%)	6 (0.4%)

three varieties were mentioned by more than half of the farmers. These were PBW 343 (21%), HD 2967 (20%) and UP 262 (12%). Fertilizer application information was captured in complete detail. This part of the survey tells names of applied fertilizers, their respective doses in splits, application time with reference to planting day and availability. Wheat crop needs to be irrigated adequately to harvest optimal yield (Zaveri and B. Lobell, 2019). In the survey, farmers were asked to provide detail information around wheat plot irrigation – availability, accessibility, number of irrigation, crop stage(s) at which irrigation applied and irrigation decisions (when to irrigate). Similarly, data was recorded around practices farmers follow to control weeds – number of times herbicide(s) applied, herbicide names, time of application with reference to planting day, number of times weeds were removed manually and time of manual operations with reference to planting day. Information on weed control measures were followed by pictorial identification of top five weeds infesting surveyed wheat plot. Weeds identified with the help of a weed poster were then ranked by farmers based on their severity of damage.

Wheat grain yields of farmers' largest wheat plot were fairly normally distributed (Figure 6). The mean value coincided at 3.0 t ha⁻¹ with standard deviation of 0.85. Twenty percent of farmers obtained yields >4 t ha⁻¹, suggesting considerable scope to increase productivity. We tried to understand gaps in production practices – why 1/3rd of farmers settled with low yields (<3 t ha⁻¹). At the end, survey recorded size of households, number of members engaged in farming, marketable surplus and percent

¹ Corresponding values represent average number of times wheat crop was irrigated with corresponding standard deviation in the parenthesis

contribution of agriculture/wheat crop in household income. Each interview ended after geo-coordinates of the surveyed plot was captured with acceptable accuracy.

Figure 5: Distribution of wheat sowing time in the survey data

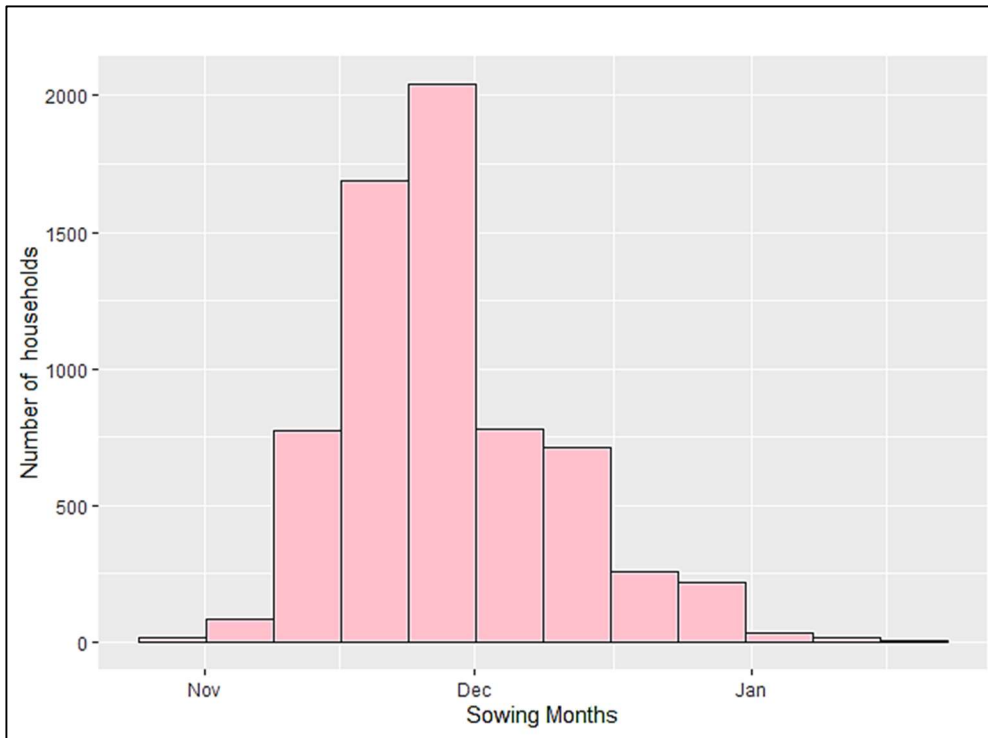
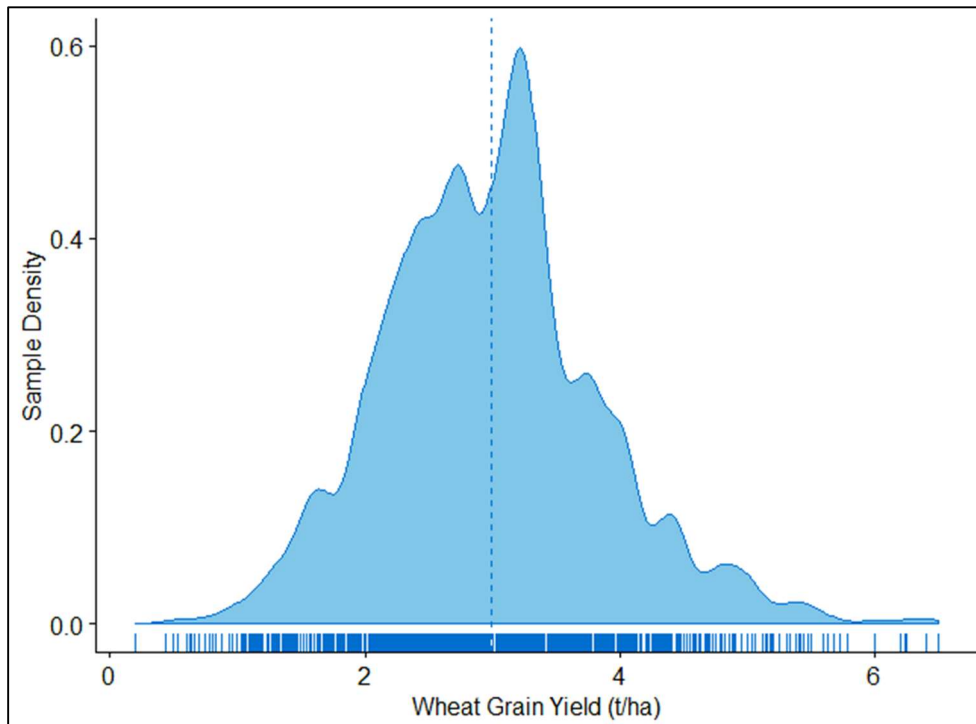


Figure 6: Distribution of wheat grain yield on the largest plots in the survey data. The vertical dashed line represents mean wheat yield (2.99 t/ha) of the sample.



4. ACKNOWLEDGEMENTS: Funding for this survey was provided by the Bill & Melinda Gates Foundation as part of the CSISA project. We sincerely acknowledge the effort put forward by our project team members (Ajay Pundir, Anurag Kumar, Pankaj Kumar, Prabhat Kumar, Deepak Kumar Singh, Madhulika Singh and Moben Ignatius) for coordinating data collection with KVK partners. We highly appreciate continued engagement of all forty KVK personnel in data collection. The authors are grateful to ICAR for collaborating in this endeavour and taking-up the survey at much larger scale through its extension wing.

REFERENCES

- A.K. Singh, Peter Craufurd, Andrew McDonald, Ajoy Kumar Singh, Anjani Kumar, Randhir Singh, Balwinder Singh, Sudhansu Singh, Virender Kumar and R.K. Malik (Eds.) 2019. New Frontiers in Agricultural Extension. International Maize and Wheat Improvement Center (CIMMYT). handle: [10883/20738](https://hdl.handle.net/10883/20738)
- Malik, RK., Singh, S., Yadav, A., 2007. Effect of sowing time on grain yield of wheat (*Triticum aestivum* L.) in rice-wheat cropping system. Haryana Agricultural University Journal of Research. 37. 103-105.
- Docs.getodk.org. 2020. [online] Available here: <https://docs.getodk.org/downloads/ODK-Documentation.pdf> [Accessed 30 August 2020].
- Sedgwick, P., 2014. Cluster sampling. *BMJ*, 348, g1215. doi: [10.1136/bmj.g1215](https://doi.org/10.1136/bmj.g1215)
- Singh, A., Malik, RK., Kumar, A., Roy B., Rajarshi & Roy, P., 2019. Comparative efficacy of improved crop management technologies of wheat (*Triticum aestivum*) in Bihar. Indian Journal of Agricultural Sciences. 89/8, 1344-1352. url: <http://epubs.icar.org.in/ejournal/index.php/IJAgS/article/view/92872>
- Skinner, C., 2020. *Probability Proportional To Size (PPS) Sampling*. doi: [10.1002/9781118445112.stat03346.pub2](https://doi.org/10.1002/9781118445112.stat03346.pub2)
- Zaveri, E. and B. Lobell, D., 2019. The role of irrigation in changing wheat yields and heat sensitivity in India. Nature Communications, 10(1). doi: [10.1038/s41467-019-12183-9](https://doi.org/10.1038/s41467-019-12183-9).