Crowdsourcing field observations from smallholder farmers in Tanzania using interactive voice response

Tim C.B. Chancellor[1*], Jan S.H. Priebe[1] and Prisila Mkenda[2]

[1] Natural Resources Institute, University of Greenwich, Chatham Maritime, UK

[2] Nelson Mandela African Institution of Science & Technology, Arusha, Tanzania

Keywords: crowdsourcing, IVR, common bean, insect pest management

* Corresponding Author: t.chancellor@gre.ac.uk

Introduction

The term crowdsourcing was first used in 2006 but the approach has been employed in various fields over a much longer period of time (Brabham, 2013). The central idea is to collect information from a large number of people in order to achieve a common goal such as to increase knowledge, solve a problem or enhance the efficiency of a process. In the realm of research crowdsourcing is closely connected with citizen science in which members of the public are engaged in a scientific project. In C17th England the naturalist John Ray arranged for a large number of volunteers to collect specimens for him and there are datasets on the phenology of plants, birds and insects in England and in some other countries which provide a continuous record for hundreds of years (Kobori *et. al*, 2016). A current example of how non-specialists are helping to gather large quantities of biological data is the Breeding Bird Survey run by the British Trust for Ornithology which engages thousands of volunteers to submit records of over 200 bird species each year (*ibid*).

Crowdsourcing in agriculture

Although crowdsourcing has been widely used in environmental monitoring there are few examples of its use in agriculture (Minet *et al*. 2017). This is surprising because there is a tradition of participatory approaches to agricultural research which have been developed in recent decades. The potential for harnessing large numbers of individual farmers to generate information on the performance of crop varieties in different locations was recently demonstrated through a triadic comparison of technologies (Van Etten *et al*., 2016). This was taken a stage further in a subsequent study in which a total of 12,409 farmers collected data on experimental plots of common bean (*Phaseolus vulgaris* L.) in Nicaragua, durum wheat (*Triticum durum* Desf.) in Ethiopia, and bread wheat (*Triticum aestivum* L.) in India (Van Etten *et al*., 2019). The results showed that expected effects of seasonal climatic variables on the performance of the crop varieties occurred and that these were generalizable across growing seasons. The authors concluded that their approach allows for more targeted recommendations for the deployment of crop varieties and that these may differ substantially from recommendations derived from current evaluation procedures in which testing is done at a limited number of sites.

Crowdsourcing has tremendous potential for use in insect pest management. Up to date information on the distribution and abundance of insect pests of crops is needed for the development of appropriate management strategies. Field surveys can be used to identify locations where pest levels are high and sequential monitoring may help to establish seasonal patterns of occurrence (Dent, 2000). Surveys are also useful in monitoring the likely sources of migrant pests and recording their dispersal from these areas. When used in combination with other information such as climatic data field surveys are an essential element of insect pest forecasting systems. In developing countries field surveys are generally conducted by agricultural research and extension staff. Their main limitation is that they are labour-intensive and costly to carry out. Since resources of manpower and funds are limited surveys are carried out infrequently and in restricted locations which may not be representative of the situation over larger areas. As a result national agricultural research and extension systems are constrained by having limited access to accurate information on the status of insect pests in different locations. This makes it difficult for them to respond to new threats such as the recent appearance and spread of the fall armyworm (*Spodoptera frugiperda*) in Africa where monitoring and early warning systems are crucial for effective control (Day *et al.*, 2017).

Data collection methods

Airborne remote sensing offers technological solutions to help address constraints caused by scarcity of labour and funds. The methods involve measuring plant spectral reflectance profiles as these are generally modified by insect herbivory (Nansen & Elliott, 2016). Satellite imagery can provide automated mapping across wide geographic areas but has several limitations for use in pest monitoring, especially in cropping systems with small fields and mixed crop cover as are commonly found in smallholder agriculture. Spatial resolution is usually low even in the most advanced systems and the quality of images can be poor when there is heavy cloud cover. Satellite system also have specialised requirements for data handling and it takes time for data to be delivered to end users. Manned aircraft and unmanned aerial vehicles (UAVs) are also used to monitor damage caused by insect pests. Although functionality of UAVs is improving rapidly and costs are reducing they are still not very practical for use in smallholder systems.

Meanwhile, less technologically advanced approaches to collecting data on insect pests through crowdsourcing have been used successfully in some locations. In the State of New Jersey in the USA Hahn and co-workers (2016) monitored the spread of an invasive insect pest, the brown marmorated stink bug *Halyomorpha halys* Stål, by crowdsourcing observations from citizens as a complement to a formal system of monitoring using a network of blacklight traps. Initially, homeowners were requested to send photographs or physical specimens via mail so that identification of the insect could be verified by experienced researchers. Subsequently, data were recorded on a form on a website which also included a facility to upload an image. Similarly, in Tanzania a community-based forecasting system was developed for the African armyworm, *Spodoptera exempta,* to contribute to a more formal national forecasting system for the insect (Holt *et al*., 2006). Residents in a network of villages recorded rainfall in rain gauges and documented catches of the insect in pheromone traps. The system was considered to be a success because it gave a greater level of agency to villagers and they had easier access to the information. It enabled the provision of location-specific early warnings of armyworm outbreaks and this led to improved control (Efa *et al*., 2010).

Utilising ICTs for information delivery

Advances in information and communication technologies have dramatically increased the capability for two-way exchange of information. Citizens in New Jersey who recorded data on *H. halys* received guidance via a website and logged their findings via email. As smartphones become more widely used the ease with which information can be shared and accessed has increased greatly. In an online survey of 57 farmers in Britain and France conducted in 2014 found that fifty-nine percent of respondents used an average of four farm-specific apps on a smartphone (Dehnen-Schmutz *et al*., 2017). A large majority of the farmers considered that citizen science was a useful methodology for collecting data as well as for other purposes. They also expressed willingness to participate in citizen science initiatives and in many cases they were prepared to spend significant amounts of time in doing so.

Mobile phone use has expanded greatly across sub-Saharan Africa but the cost of a smartphone is still beyond the means of most people, especially those in marginalised rural communities (Baumüller *et al*., 2017). Smartphone ownership among adults in sub-Saharan Africa varies by country, for example at 13 percent in Tanzania and 51 percent in South Africa, with a consistently strong bias towards more educated, higher income, younger males in urban areas (Pew, 2018). This means that basic mobile phones are likely to remain the only feasible digital way for agricultural researchers and extension staff to interact with rural households in the foreseeable future. There are many initiatives in sub-

Saharan Africa in which short message service (SMS) is used to facilitate communication with farmers (Baumüller, 2018). SMS is an extremely useful means of communication and can be used on simple mobile phones but it is not suitable for the many farmers who are not able to read and write. Interactive voice response (IVR) services are more suited to these farmers as literacy is not a requirement for their use. IVR services provide information or pose questions in the form of pre-recorded voice messages. Answers to questions are given using the numeric keys on the handset. IVR services are more flexible than SMS which has a character limit and so messages can be longer and more detailed. IVR has been used by the Kenyan National Farmers Information Service to provide extension information to farmers (Tucker & Gakuru, 2008), and Human Network International has launched the 3-2-1 information service in several African countries which utilises IVR to provide information on a range of topics, including agriculture and health (HNI, 2019).

IVR services can also be used for collecting information from farmers. However, whilst IVR services have these capabilities there is limited evidence on whether smallholder farmers in sub-Saharan Africa have the interest and capacity to utilise them to provide information about insect pests and to help manage them. We wanted to test farmers' willingness to provide information on the pest status of their fields during the growing season and to assess the feasibility of doing this through a simple system of audio messages. Such information would be useful to agricultural researchers and extension staff for monitoring the development of pest outbreaks. It might also be used to give tailored pest control recommendations to farmers in a more timely manner than is often possible at present.

With this in mind, we designed an interactive voice response (IVR) survey which we tested with a sample of farmers in the Kilimanjaro region in northern Tanzania in 2017. We were interested to explore firstly whether the way in which farmers are recruited to engage in a crowdsourcing activity affected their willingness and capacity to participate and if the participation rate reduced over time. Secondly, we wanted to establish if farmers were able to use the IVR system to respond to a simple set of closed questions and whether they would also respond to open questions about their knowledge, farming practices and information needs. Finally, we wanted to obtain some measure of the accuracy of the responses although our limited resources did not allow us to carry out a ground-truthing exercise.

Interactive voice response survey

Survey design

The survey was conducted from July to September 2016 in the Arusha region of northern Tanzania. The region has a bi-modal pattern of rainfall and common bean (*Phaseolus vulgaris* L.) is often grown as a sole crop in July-October towards the end of the long rains. The activity was linked to a project funded by the UK Darwin Initiative, with co-funding from the McKnight Foundation, which investigated the contribution of ecosystem biodiversity to *P. vulgaris* production and food security. During the project plant biodiversity surveys were undertaken across 24 farm locations in three ecological zones: low (800-1000 masl), mid (1000-1500 m) and high altitudes (1500-1800 m) with eight fields per zone in 2016 and 2017. For the IVR survey farmers were recruited via community meetings held in each of the three zones during which the nature of the study was explained, and a demonstration of the call they would receive was given and discussed in order to ensure the questions were clear. In addition to this, farmers who did not attend the community meetings but participated in a baseline survey carried out for the plant biodiversity study were recruited via automated telephone call. A small incentive was provided to all participants, irrespective of the way in which they were recruited. Each participating farmer received the sum of 10,000 Tanzanian shillings (the equivalent of approximately 4.3 US dollars) via mobile money transfer if they answered four or more phone calls.

The Darwin project focused on generating new knowledge to help farmers regulate insect pests using natural control methods and at the same time to enhance crop pollination through the provision of more favourable habitats for pollinators. Although field beans are largely self-pollinated insects do make a contribution to pollination and this was shown to be important in reducing yield loss in faba bean (*Vicia faba* L.) due to heat stress (Bishop et al., 2016). During the community meetings farmers indicated that they were especially concerned about aphid damage. They confirmed that they could readily identify aphids (*Aphis* sp.) and so it was agreed that the survey would record aphid abundance on common bean plants. Farmers indicated that they were also familiar with honey bees (*Apis mellifera scutellata*) and so these were also included in the survey.

Every seven days over a period of 11 weeks during the growing season, farmers were asked in the Kiswahili language to indicate whether:

1. Their bean field was at the seedling, vegetative, flowering, podding, or harvesting stage.
2. The number of aphids on their bean crop was less than, similar to, or more than numbers in the previous week. The same question was asked about bee populations.
3. Insect damage in the crop was less than, similar to, or more than numbers in the previous week.

4. They had used a pesticide during the week.

Responses to closed questions were made by selecting the appropriate number on the keypad of the phone. Farmers were also invited to respond to one open question, with a different question being asked each week. Examples of the questions asked are ‘Which beneficial insects do you know?’ and ‘Do you access agricultural information?’ The audio messages were sent using the Viamo IVR platform, formerly known as VOTO. After writing the survey script it was translated into Kiswahili and individual questions and messages were recorded by P. Mkenda. These recordings were uploaded to the Viamo platform where the survey was configured. The costs of using the platform were a US $1,050 subscription fee and $300 for the calls (approximately 3,240 minutes at US $0.09 per minute). A summary of the survey process is shown in Figure 1.

Participation of farmers

Three community meetings were held, one in each of the zones. A total of 112 people attended the meetings and 90 of these (80 percent) agreed to participate in survey. Sixty-three farmers were contacted via an automatic voice message and 52 of these answered the call with 45 (70 percent of those contacted) agreeing to take part in the survey. Therefore, there were 135 respondents; 42 in the high, 68 in the low and 25 in the mid zones. Overall, 67 percent of participants were female, with the group recruited by phone having a slightly higher proportion of women (71 percent vs 66 percent). The number of surveys to which farmers responded is shown in Figure 2. Response rates were broadly similar across recruitment method, zone and gender group. The only clearly observable difference was that younger farmers under the age of 35 were less likely to answer the calls. Since a small incentive was given to respondents for answering at least four calls we expected that the number of responses would reduce after the fourth call. This was not the case as the response rate remained at just over 60 percent until week 11 when it dropped to 53 percent.

Ease of use of the IVR system

Generally, farmers were quick to learn how to use the service and did not need any instructions on how to answer questions using the numeric keys on their phone. Answering open questions was slightly more challenging since this involved speaking into the phone and finishing by pressing a button. Farmers answered 35 percent of the open questions compared with 88 percent of the closed questions. Twelve percent of the responses to open questions were either unclear or unrelated to the question asked, while in 53 percent of cases no answer was given. In these cases it is not clear whether the absence of an answer was due to the fact that farmers were unsure how to provide open answers on their phones, whether they were not immediately able to think of a suitable answer or whether they were unwilling to respond. Participants who were sensitised in the community meetings were slightly more likely to provide an answer than those recruited by phone (39 percent and 26 percent, respectively). Elderly participants were least able to respond with only 25 percent of those aged 55 and above providing answers to open questions. The average time spent by farmers in answering the calls ranged from just under three minutes to almost four and a half minutes. This variation in the duration of the responses was primarily associated with the time farmers took to answer the open question that was asked in a particular week.

Farmers recruited through the automated call were equably able to respond to closed questions as farmers recruited through community meetings. On average there was no significant difference between the number of closed questions answered by the two groups over the course of the study. Since the recruitment rate through the automated call was also quite high this suggests that the method has the potential to be used to engage with large numbers of farmers at a much lower cost than through physical meetings. However, this would not be an advantage if, for example, physical meetings are needed to train farmers to correctly identify the target pest.

Reliability and accuracy of the responses

A simple check was made to assess the reliability of the responses given by farmers in one component of the data set. Any response indicating that the crop was at an earlier stage than that recorded in the previous week was considered to be incorrect. The error rate of farmers in providing this information is shown in Figure 3. The results show that, based on this criterion, nine percent of answers were erroneous and 43 percent of farmers made at least one error. There was a variation of planting dates between zones, with fields in the high zone planted later than those in the mid and low zones. The responses given by farmers showed a good fit with the pattern of crop development within each of the zones. So, whilst farmers did give some incorrect responses in the data set the overall results give a measure of confidence in their ability to provide reliable information.

Responses provided by farmers on aphid and honeybee abundance and on insect feeding damage were converted as follows: -1 for less than the previous week, 0 for no change, and +1 for more than the previous week. The figures were then cumulated over the 11-week survey period to give an ordinal scale (running from -10 to +10, as no data were available for week 10) with all responses set as zero at the start. So, for example, a succession of weekly responses showing 'no change, less than, less than, less than, more than, more than' would be logged as 0, -1, -2, -3, -2, -1. A linear Analysis of

Covariance (ANCOVA) model was applied to the data and the non-parametric Friedman test was used to assess the significance of temporal trends.

Weekly means and standard errors with fitted regression lines are shown for the three variables in each of the zones in Figure 4. As indicated by the negative scales on the y axes of the graphs, aphid and honeybee abundance declined over the survey period and insect damage reduced. These trends for data combined across the three zones were highly significant for aphid abundance (Friedman $\chi2(9) = 526.6$, $p < 0.00001$), honeybee abundance ($\chi2(9) = 345.81$, $p < 0.0001$) and insect feeding damage ($\chi2(9) = 526.58$, $p < 0.0001$). We were not able to validate the accuracy of the information provided by farmers through field observation. It might be expected, for example, that honeybee numbers would increase during the flowering stage of the bean crop towards the middle of the survey period. However, other factors such as pesticide use might have acted to prevent this. Whilst acknowledging the rather crude nature of the survey process the results do show there was good agreement between the respondents in the temporal trends. This suggests that the observations reported by farmers did at least reflect the general patterns that occurred in the field.

Conclusions

The results from our pilot crowdsourcing study showed that farmers were willing to provide different types of information, including details of an important insect pest on their crops of common bean. Concerns about privacy are sometimes listed as a constraint to gathering information in this way, particularly among commercial producers (Minet *et al*., 2017). But this probably does not apply to smallholder farmers in Africa who are less able to develop and maintain their own pest management regimes and have greater reliance on external advice. The response rate among farmers was high and they were able to use the system effectively. However, the response rate to open questions was low and further work is needed to identify the reasons for this. Ease of use of the technology is another potential barrier to participation in crowdsourcing activities and it is important to establish whether this was a factor with this system.

Another commonly expressed concern relates to the quality of data collected through crowdsourcing. Good quality data is especially important where it used in early warning systems of pest occurrence and to inform recommendations to farmers for pest control. Our findings showed that farmers did make errors in recording their answers to closed questions but the level of errors committed was relatively low and with improved guidance could be reduced further. We were not able to firmly establish the accuracy of the information provided by the farmers. However, in a recent analysis of the tricot trial approach mentioned above the authors concluded that even with a low level of reliability in their data set statistically meaningful results could be obtained from a group of less than 200 respondents (Steinke et al., 2017). This is consistent with the so-called 'Wisdom of crowds' principle which states that accurate results can be achieved from a diverse set of people when the group size is sufficiently large. Nevertheless, we suggest that a system of validation of the collected information should be used to reduce the risk that inappropriate recommendations are given to farmers. This is needed because the ability of farmers to identify insect pests varies between locations and can be limited (e.g. Abate *et al*., 2000). In order to be feasible, data validation would need to be low-cost and not involve large numbers of professional staff.

Sustaining the engagement of respondents is another potential challenge in crowdsourcing initiatives. In our survey the level of participation farmers remained high throughout most of the cropping season. They continued to provide information beyond the period during which a small incentive was offered, suggesting that they were motivated to contribute information for other reasons. It is likely that an important factor was the interest of farmers in finding improved ways to manage insect pests. When farmers recruited through the automated telephone call were asked to indicate the type of information they were most interested in receiving, pest and disease control ranked second behind production methods.

The Viamo platform provides a convenient and accessible tool to create *ad hoc* IVR services with which to gather data or provide information. The platform itself is intuitive to use and provides extensive functionality in terms of setting up calls that utilise logic and routing based on respondent characteristics or answers and timing of calls to specific groups. Costs are incurred for access to the platform (approximately US $700 per month) and for call time (US $0.09 per minute in Tanzania). Using this method to gather information from large numbers of farmers in multiple locations provides good value for money. However, in cases where IVR services are intended to be run over longer periods of time building an IVR system using an open source system such as Asterisk may be more cost effective.

We believe that the survey results are sufficiently encouraging to further develop an IVR system and to start exploring how to use the information submitted by farmers to provide improved feedback to them. This will depend on what farmers will find most useful. There are already systems in which pre-recorded advice on specific topics is available to farmers via phone calls and this could be part of the service (Baumüller, 2018). However, the added value of an IVR crowdsourcing approach would be to enable researchers and extension agents to see where pest problems are emerging and focus their efforts on these areas.

References

Abate, T., van Huis, A. and Ampofo, J.K.O (2000) Pest Management Strategies in Traditional Agriculture: An African Perspective. *Annual Review of Entomology* **45**, 631-659.

Baumüller, H. (2017) Towards smart farming? Mobile technology trends and their potential for developing country agriculture. In Handbook on ICT in Developing Countries, Skouby KE, Williams I, Gyamfi A (eds). River Publishers: Delft.

Baumüller, H. (2018) The little we know: an exploratory literature review on the utility of mobile phone enabled services for smallholder farmers. *Journal of International Development* **30:** 134–154.

Bishop, J., Jones, H.A., Lukac, M. and Potts, S.G. (2016) Insect pollination reduces yield loss following heat stress in faba bean (*Vicia faba* L.). *Agriculture, Ecosystems & Environment* **220**, 89-96.

Brabham, D.C. (2013) *Crowdsourcing*. The MIT Press. 168pp. ISBN: 9780262518475

Day, R., Abrahams, P., , M., Beale, T., Clottey, V., Cock, M., Colmenarez, Y., Corniani, N., Early, R., Godwin, J., Gomez, J., Moreno, P.G., Murphy, S.T., Oppong-Mensah, B., Phiri, N., Pratt, C., Silvestri, S. and Witt, A. (2017) Fall armyworm: impacts and implications for Africa. *Outlooks on Pest Management* **28**(5), 196-201.

Dehnen-Schmutz, K., Foster, G.L., Owen, L. and Persello, S. (2017) Exploring the role of smartphone technology for citizen science in agriculture. *Agronomy for Sustainable Development* **36**: 25. https://doi.org/10.1007/s13593-016-0359-9

Dent, D. (2000) Insect Pest Management. 2nd edition. CABI International, Wallingford, UK.

Hahn, N.G., Kaufman, A.J., Rodriguez-Saona, C., Nielsen, A.L., Laforest, J. and Hamilton G.C. (2016) Exploring the spread of brown marmorated stink bug in New Jersey through the use of crowdsourced reports. *American Entomologist* **62**(1), 36-45.

Efa, N., Musebe, R., Day, R., Romney, D., Kimani, M., Maulana, T. and Mallya, G. (2010) Integrating indigenous and exogenous communication channels and capabilities through community-based armyworm forecasting. *African Journal of Crop Science* **18**(3), 115-125.

Holt, J., Mushobozi, W., Day, R.K., Knight, J.D., Kimani, M., Njuki, J. and Musebe, R. (2006) A simple Bayesian network to interpret the accuracy of armyworm outbreak forecasts. *Annals of Applied Biology* **148**, 141–146.

Human Network International (HNI) (2019) A search engine where there is no Internet. Retrieved from http://hni.org/what-we-do/3-2-1-service/

GSMA (2017) The mobile economy 2018. Report downloaded on 15 April 2019 at https://www.gsma.com/mobileeconomy/wp-content/uploads/2018/05/The-Mobile-Economy-2018.pdf

Minet, J., Curnel, Y., Gobin, A., Goffart, J-P., Mélard, F., Tychon, B., Wellens, J. and Defourny, P. (2017) Crowdsourcing for agricultural applications: a review of uses and opportunities for a farmsourcing approach. *Computers and Electronics in Agriculture* **142**, 126-138.

Nansen, C. and Elliot, N.C. (2016) Remote sensing and reflectance profiling in entomology. *Annual Review of Entomology* **61**:139–158.

Pew Research Center (2018) Internet connectivity seen as having positive impact on life in sub-Saharan Africa. Report downloaded on 17 April 2019 at https://www.pewglobal.org/wp-content/uploads/sites/2/2018/10/Pew-Research-Center_Technology-use-in-Sub-Saharan-Africa_2018-10-09.pdf

Steinke, J., van Etten, J. and Zelan, P.M. (2017) The accuracy of farmer-generated data in an agricultural citizen science methodology. *Agronomy for Sustainable Development* **37**, 32.

Tucker, R.C.F. and Gakuru, M. (2008) Experience with developing and deploying an agricultural information system using spoken language technology in Kenya. In: Proceedings of the IEEE Spoken Language Technology Workshop (15-19 December 2008), Goa, India (pp. 17-20). Retrieved 23 April, 2019, from IEEE Xplore.

Van Etten, J., Beza, E., Calderer, L., Van Duijvendijk, K., Fadda, C., Fantahun, B., Kidane, Y.G., Van de Gevel, J., Gupta, A., Mengistu, D.K., Kiambi, D., Mathur, P.N., Mercado, L., Mittra, S., Mollel, M.J., Rosas, J.C., wals, J., Suchini, J.G. and Zimmerer, K.S. (2016) First experiences with a novel farmer citizen science approach: crowdsourcing participatory variety selection through on-farm triadic comparisons of technologies (TRICOT). *Experimental Agriculture*, 1-22. doi:10.1017/S0014479716000739

Van Etten, J. et al. (2019) Crop variety management for climate adaptation supported by citizen science. *Proceedings of the National Academy of Sciences* **16**(10), 4194–4199. https://www.pnas.org/cgi/doi/10.1073/pnas.1813720116

Acknowledgements

We would like to thank the team of the Darwin Initiative project on 'Harnessing agricultural ecosystem biodiversity for bean production and food security' for logistical support and advice. We extend our appreciation to the McKnight Foundation for providing funds to conduct the study through its Collaborative Crop Research Program. We thank Dr Stephen Young for his input to the statistical analysis of the survey data. We are especially grateful to the agricultural extension staff and farmers in the Arusha region who gave their time to take part in the research.