Learnings from Project FRAME

Introduction

FRAME (Fostering Resilience in Agriculture through MRV Experimentation) was implemented in Bihar, India, from 2021 to 2023 to enhance climate resilience in crop production for small-scale farmers. The project was led by Digital Green (DG) and supported by partners Tata Cornell Institute (TCI), the Environmental Defense Fund (EDF) as technical collaborators and the Bihar Rural Livelihood Promotion Society (JEEViKA) as prime implementation partner.

The project focused on raising awareness of sustainable agricultural practices and promoting data-driven decision-making. The project collected datasets of over 24,000 farmers which were processed to feed into a digital tool called Farmer Scorecard that advanced data-based decision making in smallholder fields (averaging 0.34 Ha) by providing plot-specific advice, optimized resource use (mainly fertilizers), and improved farming techniques, particularly for rice and wheat cultivators. This combination of technology and sustainable approaches like the optimization of resources (fertilizer application) has been instrumental in maximizing the program's impact on farmer's lives. These innovations are also informing follow-up initiatives like the Zero Hunger Zero Carbon Project.

The project was co-funded by Sequoia Climate Fund, CISCO Foundation, and the initiative of German Development Cooperation FAIR Forward - Artificial Intelligence for All, with Zindi Africa hosting the Crop Yield Estimation Data Challenge.[1]

Key Objectives:

(1) One of the project's objectives was to develop a thorough **open-source dataset with smallholder farmers on their agricultural practices in Bihar**. This dataset was compiled and managed through in-person farmer surveys, utilizing digital tools to capture detailed information on agricultural management practices, crop yields, and residue management.

(2) This dataset served as the foundation for a **Digital Green Crop Yield Estimation Challenge that aimed to create a machine learning solution to predict the crop yield per acre of rice & wheat crops** in Bihar, India.

This report presents the insights and lessons learned in accordance with our commitment to create a data-sharing ecosystem, where we priorite the consent and benefits of the farmers who contribute to the data, ensuring that they directly benefit from any technological advancements derived from the crop yield estimation models.

---

[1] Zindi. "Digital Green Crop Yield Estimate Challenge." Zindi Africa. Accessed August 8, 2024. https://zindi.africa/competitions/digital-green-crop-yield-estimate-challenge.

Key Achievements

1: Data Collection & Management:

The strategy for managing the agricultural datasets centered on **creating a scalable data collection model** that could be easily integrated into the ongoing responsibilities of Jeevika's extension agents. To achieve this, the data collection processes of these agents were studied and analyzed, and areas were identified where checks and balances could ensure accurate data collection without adding to the existing workload. Simple changes, such as implementing data validation for fertigation requirements and land under cultivation, helped reduce the probability of errors. The language of the forms was also simplified, making them comprehensive and self-explanatory. Additionally, the available data collection tools were reviewed, and KOBO was shortlisted due to its offline functionality, which was ideal for the identified areas with low internet connectivity.

These forms and processes were beta tested before scaling, and 275 extension agents were trained to use them and ask pertinent questions from farmers on their agricultural practice throughout the crop cycle. Through these agents, agronomic data was collected and processed to create farmer scorecards. Data was gathered from 14,126 farmers for the wheat crop in the winter of 2022-23, and from 24,140 farmers for the rice crop during the monsoon season of 2023.

TCI helped finalize the attributes required for data collection, such as nursery management, including seed selection, timing, the use of fertilizer or manure, and the age of seedlings during transplantation. Among the several land plots available with the identified farmers, data was collected from the largest ones to ensure that practice and other kinds of data that was collected was significant. By focusing on the largest plot, the survey aims to capture the most representative, accurate and relevant data for decision making at both individual farm level and for broader agricultural planning. These land plots were also geofenced.

For rice the data points collected focused on nursery management, including aspects such as seed selection, timing, the use of fertilizer or manure, and the age of seedlings during transplantation. For both rice and wheat, land management details were recorded, encompassing the type of land, number of tillage operations, timing, basal dose of fertilizers, and irrigation methods employed. Nutrition management data included information on fertilizer top dressing events. Additionally, irrigation management details were documented, such as the number of irrigation events, timing, and the land's water-holding capacity. Major pest attacks and diseases affecting the crops were also noted. Lastly, data on yield and residue management practices were gathered.

The data is available as open source, and the link to access it is

https://github.com/digitalgreenorg/frame-templates

Consent was a fundamental aspect of the data collection process. **Farmers were fully informed about the purpose of gathering their data, its intended use, and the potential benefits they could expect**. The FLEWs received thorough training, not just on the data collection tools, but also on the importance of the data collection and how it would assist farmers in making critical decisions regarding their farming practices. Each data collection form began with a consent statement, which the FLEWs read aloud to the farmers, explaining the purpose and data privacy terms. The process continued only if the farmer provided explicit consent; otherwise, the survey was terminated immediately.

The data collection model, through FRAME, was refined to achieve greater efficiency, with an e**xemplary data validation protocol established for rigorous data quality assessment (Annexure 1)**. The range for the amount of Urea/Diammonium Phosphate (DAP), sowing/transplantation dates, and the number of irrigations was determined as a one-time activity for the season. The range was determined using the recommendations of the TCI team relating the range to yield, serving as an alternative to soil quality tests for determining fertilizer needs. This approach was developed from TCI's field trials and the recommendations generated for farmers. Data was analyzed weekly (for the protocol, please see Annexure 1) to determine block-wise averages and ranges for DAP/Urea usage, sowing dates, and irrigation. Additionally, correlations between Urea/DAP and other factors are examined. The next step involved weekly data validation, reliability, and consistency checks. The state program team validated 10% of outliers, shared insights with JEEViKA block officers, and conducted reliability and consistency tests. Learnings were then compiled and shared with the team to ensure continuous improvement and effective dissemination of information.

Highlights

1. The Frontline extension workers are salaried employees[2] of the Bihar Rural Livelihood Promotion Society (JEEViKA) and are required to perform various tasks including providing training, field inspections, demonstrations and farmer field schools. For the collection of agricultural datasets required for this project, the team had to design a scalable and integrated data collection model that fit seamlessly in the existing workflows without increasing it.
2. Key improvements included simplifying and digitizing data collection forms, enhancing data validation processes, and using tools like KOBO for areas with limited internet connectivity.
3. A strong focus was put on the training and onboarding of these FLWs to ensure they could effectively use the digital tools and understand their purpose
4. Geofencing and focusing on the largest land plots ensured more accurate and representative data.

---

[2] Bihar Rural Livelihoods Promotion Society. "Village Resource Person (VRP) and Social, Economic, and Wellbeing (SEW) Policy Paper." Bihar Rural Livelihoods Promotion Society, Government of Bihar, July 2018. https://www.brlps.in/UplodFiles/Order/VRP%20and%20SEW%20Policy%20Paper.pdf.

5. Consent was prioritized to maintain transparency and data privacy.
6. The refined data collection model, including rigorous validation protocols and weekly data analysis, demonstrated the value of continuous improvement and effective communication for better decision-making and agricultural planning.

2: AI Data Challenge:

Zindi Africa was engaged to host and present the awards for the data challenge. The aim of the challenge was to **create a machine learning solution to predict the crop yield per acre of rice or wheat crops in Bihar, India**. A crop yield model has the potential to improve Indian agriculture and can serve as a benchmark for smallholder farmers. By providing near accurate yield predictions, the model enables farmers to make informed decisions about planting and resource allocation, which can help reduce poverty, combat malnutrition, and enhance food security. As climate change accelerates, adaptive farming practices become increasingly important, making precise yield predictions even more essential. Innovations in this area can drive sustainable agriculture and support a stable food supply for the expanding population.

The insights gained from the data challenge have practical applications in real-world scenarios, such as enhancing agricultural practices, optimizing resource allocation, and enabling informed decision-making for farmers, agricultural policymakers, and stakeholders. These yield prediction models are especially valuable for farmer organizations as they can estimate future production levels when negotiating with traders. Additionally, the models offer guidance to farmers on enhancing their practices, which can lead to better quality and greater quantities of produce. Successful practices can be promoted by government policymakers, who can incentivize farmers to adopt these improved farming practices.

Although the models performed well with small, practice-level datasets focused on rice, its effectiveness seems to be less productive when applied to larger datasets involving different crops, such as wheat. However, we are optimistic about testing it on the recently collected rice data from the 2023 monsoon season and will share updates on the results of this analysis.

The possible scalable model of The Oxford Effective International Development Group, one of the winners in the data challenge, was further analyzed using the data provided, to conduct pseudo-causal studies on the impact of various fertilizer application methods on crop yields. Their initial findings were highly encouraging, leading them to seek publication in academic circles. As a result, they approached DG for additional data to validate their results, and DG anticipates a collaborative publication with the group.

The winning solutions are available in open source and can be access using the following link https://github.com/digitalgreenorg/frame-templates

Highlights

1. The data challenge event underscored the crucial role of machine learning models in advancing farming practices for farmers in Bihar, India.

2. The winning models hold promise for enhancing agricultural techniques, aiding farmers in making informed decisions about resource use. Such decisions can lead to better resource management, improved production efficiency, and strengthened food security, especially in the face of climate change's growing demands for adaptive practices. The challenge's findings have practical implications for optimizing resource allocation and assisting farmer organizations in negotiations. However, while the models demonstrated promising results with smaller rice-focused datasets, their performance with larger, more diverse datasets remains uncertain. The projected yield obtained from the model needs to be triangulated with actual yield across different agro-climatic zones across years to authenticate the accuracy of the model. Future analysis of recent rice data is anticipated to offer additional insights.

3: Climate Action & Farmer Awareness:

During the project duration, Digital Green, in consultation with TCI, developed and disseminated audio-visual advisories on climate-smart farming practices to address gaps in farmers' information on the right practice. A total of 8 video advisories were produced for both paddy and wheat. These included 3 introductory videos on the correlation between climate and agriculture and the advantages of Climate-Smart Agriculture (CSA). Specific advisories included the benefits of early transplantation of rice, and fertilizer optimization based on the average yield of the last 3 years. For wheat, videos covered the advantages of early sowing, and fertilizer optimization based on the number of irrigations and sowing time. Additional videos discussed irrigation management in wheat, highlighting critical stages of irrigation.[3] The video advisories were distributed to over 300,000 farmers through 275 FLEWs across 11 districts, spanning 3 agroclimatic zones of Bihar.

Farmer Scorecards

DG's Farmer Scorecard developed on the basis of the practice level data of the farmer to disseminate targeted, plot-specific advisories, comparing individual farmers to their peers and recommended dosages/practices. It not only provided farmers with visibility of their own practices such as quantity of fertilizer applied but also prompted them to adopt advantageous practices based on social proof. Using the recommendation from the scorecard, farmers could apply exact quantities of fertilizers, rather than relying on generic recommendations. The usage data of the farmers was color coded. Green data indicated correct usage and red data indicated that the usage was incorrect and needed improvement.

---

[3] (Playlist - https://youtube.com/playlist?list=PL-WsPllTgj_4qplgiVV_eONu6YOEN-IlC&si=-CPFb3pqfqJ7h9h4)

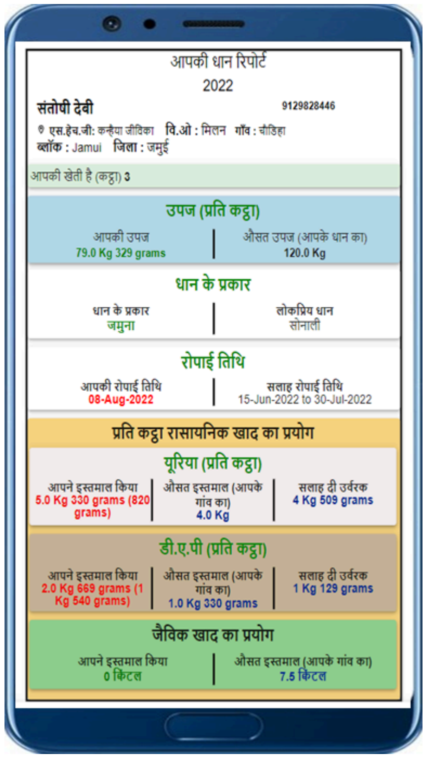

Fig 1: Farmer Scorecard

About 4,123 farmers used farmer scorecards in the pilot phase from a total of 24,140 rice farmers in 2023. Data for these farmers was collected over two consecutive years, during the monsoon seasons of 2022 and 2023. The analysis revealed a significant increase in early transplantation among the farmers, indicating that the communication of these recommendations was effective. In 2023, the percentage of farmers opting for early transplanting (in July or earlier) significantly increased, rising from 71.1% in 2022 to 92% in 2023 among the 4,123 farmers whose data was collected during both monsoon seasons. A similar trend was observed among the remaining 19,930 farmers of monsoon 2023, who did not receive a scorecard but were provided with video advisories. Consequently, the percentage of farmers transplanting late (after July) decreased from 28.9% in 2022 to just 8% in 2023. The adoption of practices in one crop often leads to the adoption of practices in the subsequent crop, thereby increasing the benefits to the farmer throughout the cropping cycle to utilize the entire season for multi-cropping and possibility of better return.

Additionally, there was better adherence to nitrogen use recommendations among those who received both the scorecards and video advisories. In the "**Scorecard + Video Advisory**" group (n = 4,123), a higher proportion of farmers (27.60%) used nitrogen close to the recommended level, compared to the "Only Video Advisory" group (n = 19,930), where fewer farmers (15.75%) did so. Overall, the "Scorecard + Video Advisory" intervention seems effective in helping farmers get closer to the recommended nitrogen use, as it had a higher percentage of farmers in the near recommended category. The "Only Video Advisory'' intervention shows that more farmers used less than the recommended nitrogen, suggesting that additional support, like scorecards, may help align nitrogen use better with recommendations.

Highlights

1. The project significantly improved farmers' understanding of climate-smart agriculture (CSA) by developing and sharing targeted audio-visual advisories. It produced eight instructional videos focusing on sensitizing farmers on climate change and its impact on agriculture (3), paddy (2) and wheat (3) cultivation, addressing key practices such as early transplantation, fertilizer optimization, and irrigation management, all tailored to various agro climatic zones in Bihar.
2. Through the sensitization videos, the farmers were made aware of the changing climatic risks and then the practice videos helped them to adopt the recommended practices.
3. Farmer Scorecards provided personalized feedback by comparing individual practices with peer benchmarks, promoting accurate fertilizer dosage and highlighting areas for improvement. This strategy not only spread essential information but also encouraged the adoption of best practices through both visual guidance and peer influence

4: **Ecosystem Support**

Digital Green hosted and participated in key climate and agriculture events, sharing insights on their MRV (Measurement, Reporting, and Verification) approach.

a. Bihar Summit 2023 on Climate-Resilient Livelihoods
b. FRAME Stakeholders Meeting including international non-governmental organizations (INGOs), carbon project developers, multilateral agencies, research institutions, government bodies, and development agencies.
c. International Rice Congress 2023 (a global scientific platform for rice-based food systems innovators to work towards shaping a food- and nutrition-secure future), held in October 2023 in the Philippines
d. Asia Pacific regional meeting of the " 4 per 1000 Initiative (A global initiative aiming to show that agriculture can provide concrete solutions to the challenge posed by climate change) in September 2023
e. A stakeholder consultation with the Environmental Defence Fund resulted in a report titled "Improving Fertilizer Use Efficiency for Farmers, Communities, and the Climate." The report highlights that only 40% of nitrogen in fertilizers is absorbed by crops, with the rest lost as greenhouse gasses, leading to environmental harm. It underscores the need to address economic and behavioral barriers to adopting sustainable nitrogen use practices and references government initiatives like PM PRANAM that support improved fertilizer efficiency.

Summary and Future Work

In summary through this project:

1. A scalable and integrated data collection model was developed for JEEViKA's frontline workers, incorporating digital tools, geofencing, and rigorous validation while prioritizing training, consent, and continuous improvement to enhance agricultural planning and decision-making.
2. The data challenge highlighted the potential of machine learning models to improve farming practices and resource management in Bihar, India, though further validation with larger, diverse datasets and actual yields is necessary for broader application.
3. The project enhanced farmers' understanding of climate-smart agriculture in Bihar by using targeted videos and Farmer Scorecards to raise awareness of climate risks, promote best practices, and provide personalized feedback for improved farming techniques.

Building on the insights gained from the FRAME project, the upcoming Zero Hunger Zero Carbon (ZHZC) project will focus on refining and scaling data collection methodologies and enhancing digital advisory services for farmers through multiple modalities (IVR, SMS, WhatsApp, Telegram), in different languages/dialects in voice, text, and images. The primary goal remains to strengthen farmers' resilience to climate change while ensuring that data collection practices are ethical, beneficial, and scalable. Recognizing the challenges in gathering large datasets, the ZHZC project will also explore alternative methods, such as self-reporting by farmers on their practices and adoption rates.

Furthermore, the project will expand its advisory services to include livestock management, offering a holistic approach to improving the farming community's economy. These advisories will cover the relationship between livestock and climate, ration management for large ruminants, and the advantages of using sex-sorted semen for artificial insemination to boost milk productivity and managing productive herd size. The project also includes piloting on agri-voltaics to expose the communities on regenerative energy sources and how this can be integrated with crop production energy needs.

The FRAME project's findings highlighted the significant potential of farmer scorecards in increasing adoption rates. To build on this success, there are plans to make the digital farmer scorecard more comprehensive, incorporating advisories on various precision farming practices beyond just fertilizer optimization and crop timing.

Building on the success of the data challenge, Digital Green will actively seek opportunities to collaborate with partners and academic institutions to explore meaningful applications of the collected datasets, both current and future. The organization will continue to present findings and success stories from the project in relevant forums, further advancing the initiative's impact.

Annexure 1: Data Validation and Quality Assessment Protocol

The data validation process created for the project comprised the following activities:

Step 1 Involved defining the range for the quantity of Urea/DAP, date of sowing/transplantation, and number of irrigation events as a one-time activity.

Step 2 carried out weekly, this step focused on data cleaning and identifying outliers in Urea/DAP amount, sowing dates, irrigation numbers, and data duplication.

In **Step 3** Data was analyzed weekly to determine block-wise averages and ranges for DAP/Urea usage, sowing dates, and irrigation, also checking correlations between Urea/DAP use and video disseminations.

Step 4 involved data validation, reliability, and consistency checks weekly with the state program team validating 10% of outliers, sharing insights with partner officials, and conducting reliability and consistency tests. Finally, compilation of learnings is done and shared with the rest of the team.

Step 5 conducted monthly, involved representing data trends and statistical analysis.

Finally, **Step 6** included identifying frontline workers with the most outliers and providing refresher training monthly or bi-monthly.