Global Food Security 23 (2019) 256–266

Contents lists available at ScienceDirect

Global Food Security

journal homepage: www.elsevier.com/locate/gfs

A scalable scheme to implement data-driven agriculture for small-scale farmers

Daniel Jiménez[a,*], Sylvain Delerce[a], Hugo Dorado[a], James Cock[a], Luis Armando Muñoz[a], Alejandro Agamez[b], Andy Jarvis[a]

[a] International Center for Tropical Agriculture (CIAT), Cali, Colombia
[b] National Cereals and Legumes Federation (FENALCE), Colombia

ARTICLE INFO

Keywords:
Data-driven
Maize
Machine learning
Small-scale farmers
Colombia
Collaboration

ABSTRACT

The Colombian Ministry of Agriculture Colombia, an international research center and a national farmers' organization developed a data-driven agricultural program that: (i) compiles information from multiple sources; (ii) interprets that data; and (iii) presents the knowledge to farmers through the local advisory services. Data was collected from multiple sources, including small-scale farmers. Machine learning algorithms combined with expert opinion defined how variation in weather, soils and management practices interact and affect maize yield of small-scale farmers. This knowledge was then used to provide guidelines on management practices likely to produce high, stable yields. The effectiveness of the practices was confirmed in on-farm trials. The principles established can be applied to rainfed crops produced by small-scale farmers to better manage their crops with less risk of failure.

1. Introduction

Climate variability accounts for 39% of the year to year yield variability of maize (Ray et al., 2015). Mitigating the effects of this climatic variation is an important element for food security where maize is a food staple. Eighty two percent of the global maize production area is rainfed (Rosegrant et al., 2002), and outside the major maize growing areas of the world, most of the maize is produced by small-scale farmers. Around one-quarter of the world's food is produced on farms under 2 ha (Herrero et al., 2017; Ricciardi et al., 2018). Small-scale farms in developing countries often face severe financial and infrastructural constraints, and only a small fraction have access to new commercially available advances in digital agricultural technologies, and most farms remain without internet access (Mehrabi et al., 2018). Furthermore, small-scale farmers in developing countries neither routinely keep farm records nor do they have ready access to information on the weather conditions (Howland et al., 2015). Hence, they are not normally able to analyze what happened in the past and make data-based decisions for the future. Within this context, many small-scale farmers obtain knowledge on which crops to grow and how to manage them from other farmers and from extension services (Howland et al., 2015; Kiptot and Franzel, 2015; Landini, 2016). This farmer to farmer exchange of information is a cost effective alternative to conventional farmer training approaches (Nakano et al., 2018).

Farmers supported by strong growers' organizations with well-established research programs are frequently given a take it or leave it technological package based on results from experimental plots managed by researchers (Lacy, 2011; O'Neil, 2016; Steinke et al., 2017). Nevertheless, farmers, in general, prefer to discuss recommendations and make their own decisions, suited to their particular conditions, rather than being told what to do (Ingram, 2008). The packages, transmitted by extension agents, are often top-down with little opportunity for farmers to discuss how they should be adapted to their specific conditions in both space and time (Landini, 2016; Rosenheim and Gratton, 2017; Van Asten et al., 2011). There is an opportunity to use modern data management and communications technologies to provide small-scale farmers with data-driven guidelines to better manage their crops and, hence, to reduce the large year to year yield variation and increase food security.

We used the case of maize (*Zea mays*) in Cordoba, Colombia to develop a data-driven agricultural methodology in which farmers contribute information on their own experiences and combine this with information from multiple sources to provide the basis for improved crop management decisions. The basic premise is that farmers or growers are constantly producing crops under a wide range of management practices and varied growing conditions, and that a structured

* Corresponding author. Km 17 Recta Cali-Palmira, 763537 - A.A. 6713, Cali, Colombia.
E-mail address: d.jimenez@cgiar.org (D. Jiménez).

https://doi.org/10.1016/j.gfs.2019.08.004
Received 3 May 2019; Received in revised form 16 August 2019; Accepted 21 August 2019

interrogation of their observations can lead to improved management (Cook et al., 2013; Cock et al., 2011). We believe this approach, which we describe in this paper, can be readily adapted and applied to other crops and other regions of the world. This paper describes how we developed the data-driven program based on a coordinated multi-disciplinary and multi-institutional team. We hope that our experiences, successes and failures, which we present as a case study, will be helpful for others who wish to provide data-driven insights that help farmers make better decisions on crop management that will increase food security in the context of climatic variation.

Farmers in Colombia were concerned about their crops losses in 2010–2011 due to cold phases of El Niño Southern Oscillation (ENSO), also known as "la Niña", which are associated with greater than normal precipitation in certain parts of the country. Economic losses were an estimated US $7.8 billion in the country in 2010 (Esquivel et al., 2018; Hoyos et al., 2013). Countrywide, growers' associations, including the National Cereals and Legumes Federation (FENALCE), asked the Ministry of Agriculture and Rural Development of Colombia (MADR) to help them deal better with extreme weather effects. The International Center for Tropical Agriculture (CIAT) was developing new means of using information from multiple sources, including the growers, to help farmers make better decisions suited to their local conditions. At that time, a former minister of agriculture, who was also a member of the CIAT board of directors, saw the opportunity to bring various agencies together to face their common concerns. In 2013, after a series of meetings promoted by the former minister of agriculture, various growers' federations, including FENALCE, and CIAT signed an umbrella agreement with the MADR to strengthen the Colombian agricultural sector with emphasis on the capacity to adapt to both weather variability and climate change. (We use the term weather for shorter term variation in meteorological conditions according to the axiom of Robert A. Heinlein "Farmers get weather, not climate". We reserve climate for averages of precipitation and other measures of the weather that occur over a long period in a particular place). Distinct roles were defined for the agencies involved. The MADR role was to provide funding, determine priority crops and regions and designate major responsibilities. Within this framework, CIAT would provide scientific knowledge, propose and develop technological digital agricultural tools, including a web-based system, machine learning algorithms and training the staff of collaborating agencies in the use of these tools.

The MADR decided that maize in Cordoba was high priority, which led to the establishment of the Cordoba Maize Program described here. The MADR invited FENALCE, a member of the original umbrella agreement, to work with CIAT to develop a program to support maize growers in Cordoba in their efforts to obtain not only higher, but critically, more stable yields.

FENALCE and CIAT decided that the focus would not be the traditional one, with a research to extension service to farmer approach (Bänziger et al., 2006; Evans and Fischer, 1999; O'Neil, 2016). Rather, the idea was to: (i) bring together information from multiple sources; (ii) analyze and interpret the information; and (iii) present the knowledge generated to decision makers at all levels in a format that they could both understand and use. This approach complements traditional research results by integrating information from multiple sources including those of the farmers themselves (Araya et al., 2010; Lacy, 2011). Hence, the use of observational data to help farmers make better decisions in no way negates the need for research based on carefully controlled experiments: rather the two approaches are complementary. Thus, for example, in modern agriculture farmers with observational data will not produce a new cultivar, however, observational data can provide major insights on how to manage a new variety once it is produced (Cock et al., 2011; Sagarin and Pauchard, 2010). Furthermore, this method provides technology adapted to local conditions: productivity is linked to practices, the weather and the characteristics of individual farmers' fields at a precise moment in time. Aggregated data would not provide the detail needed: Information is required from individual harvests and how they were produced.

Multiple institutions and agencies with distinct roles and expertise would have to collaborate to obtain data from individual farms over several years, analyze the data and provide actionable information to farmers and other decision makers. The roles of the MADR and CIAT continued to be those foreseen in the umbrella agreement. The role of FENALCE was to bridge the gap between scientists and practitioners; to facilitate communication with farmers; to learn how to use the digital tools and methods and make them available to others; and to interpret and share the results of analyses with a view to helping farmers make better management decisions.

Early in the discussions between MADR, CIAT and FENALCE, two aspects of climatic variation were distinguished: stochastic variation in weather patterns and long-term trends in climate change. The MADR identified stochastic variation as the immediate problem, while being aware that the approaches could be adapted for long-term climate change trends. Thus, the program emphasized the short-term stochastic variation without forgetting longer-term trends (CIAT-MADR, 2015; Esquivel et al., 2018; Loboguerrero et al., 2018). The aim was to obtain as much information as possible from farmers and associate the variation in yield with management, weather, and soil and topographic characteristics. From these associations, improved management options tailored to the specific weather, soil and topographic conditions of a given field or farm at a specific moment in time can be deduced (Cock et al., 2011). As a first approximation we concentrated on the variation at the field level rather than that within fields. (Cock et al., 2016; Jiménez et al., 2016).

2. Methodology

The program was based on: the collection and compiling of basic data on the crop and the factors that influence productivity; analysis and interpretation of the data collected; and presentation of the knowledge generated to the growers and their representatives in a readily understandable and useable format.

2.1. Data collection and cleaning

Records kept by farmers or farmers' organizations were a possible source of data. FENALCE indicated that few farmers kept records. FENALCE maintained aggregated information on estimates of area planted and harvested, with estimates of mean yields and costs, but no data on soils and weather. FENALCE decided to set up a modern web-based system to collect and maintain data from individual farms. The web-platform (later named SIRIA) would be the backbone of the whole information system. The overall structure was designed to fit the needs of FENALCE, but with a view to it being easily scaled to other similar organization with other crops. Details of user profiles and management are described in supplementary information section (S1.1). Variables to be collected on the cropping events were chosen in an iterative process with CIAT and FENALCE looking for a balance between precision, and the time required to complete the forms for each cropping event (Table S1). The criteria for the overall collection of data were that it would be sufficiently detailed to characterize the environment and the management of individual cropping events in such a manner that associations between crop performance, the environment and crop management could be established. Expert knowledge of the crop provided by agronomists guided this process. A cropping event occurs in a site within a given period (Cock et al., 2011). For maize this period is taken as planting to harvest and the site is taken as an individual field. Details of the variables collected are given in the supplementary information. (S1.2).

In order to obtain soil characteristics of individual fields, we first considered using the layers from the national geographic institute Augustin Codazzi (https://www.igac.gov.co/). These were rejected as the 40 km spatial resolution was too coarse. An *in situ* Rapid Soil and

Table 1
List of the variables integrated into the models.

Variable name/predictor	Meaning	Type	Unit
Sowing_Method	Sowing method	Categorical	
Sowing_Seeds_Number	Seeding rate	Discrete	Freq.
Seeds_Treatment	Seeds Treatment (Yes/No)	Categorical	
Seeds_Per_Site	Number of seed per site	Discrete	Freq.
Cultivar	Name of cultivar	Categorical	
Former_Crop	Former crop	Categorical	
Field_Drainage	Field Drainage (Yes/No)	Categorical	
Plant_Density_20_days	Plan density in 20 day after the sowing date	Continuous	plants/mts^2
Harvest_Method	Harvest method	Categorical	
Cultivar_Type	Cultivar type	Categorical	
Chemical_Treat_Disease	Number of chemical controls to treat diseases	Discrete	Freq.
Chemical_Treat_Weeds	Number of chemical controls to treat weeds	Discrete	Freq.
Chemical_Treat_Pests	Number of chemical controls to treat pests	Discrete	Freq.
Total_N	Amount of nitrogenous applied	Continuous	t/ha
Total_P	Amount of phosphorus applied	Continuous	t/ha
Total_K	Amount of potassium applied	Continuous	t/ha
Number_Chemical_Ferti	Number of chemical fertilizations	Discrete	Freq.
pH	pH	Continuous	
Soil_Structure	Soil Structure	Categorical	
Efective_Depth	Effective soil depth	Continuous	cm
Runoff	Run-off	Categorical	
Soil_Texture	Soil Texture	Categorical	
Organic_Matter_Content	Organic Matter Content	Continuous	%
TM_Avg_VEG	Average minimum temperature.	Continuous	°C
TA_Avg_VEG	Average temperature.	Continuous	°C
TA_Avg_CF	Calculated separately on each growth stage		
DR_Avg_VEG	Average diurnal range.	Continuous	°C
DR_Avg_CF	Calculated separately on each growth stage		
DR_Avg_MAT			
TX_34_Freq_MAT	frequency of days with maximum temperature above 34 °C:	Continuous	N.A.
P_Accu_VEG	Accumulated precipitation.	Continuous	mm
P_Accu_CF	Calculated separately on each growth stage		
P_10_Freq_VEG	Frequency of days with more than 10 mm precipitation (considered as significant rainfall)	Continuous	N.A.
P_10_Freq_MAT	Calculated separately on each growth stage		
RH_Avg_MAT	Average relative humidity	Continuous	%
RH_Avg_CF	Calculated separately on each growth stage		
SR_Accu_VEG	Accumulated solar energy.	Continuous	Cal. cm-2
SR_Accu_MAT	Calculated separately on each growth stage		
SR_Accu_CF			
Yield	Crop productivity	Continuous	t/ha

Terrain Assessment (RASTA), available on-line at https://cgspace.cgiar.org/handle/10568/69682 was used. RASTA allows a farmer or technician to describe a soil and the topography in the field (Cock et al., 2002).

Previous experiences with data collected on Paper Forms (PF) and then digitalized highlighted the problem of increased errors and delays in data being available for analysis (Jiménez, 2013). Looking to the future we decided to co-design with FENALCE a smart-phone application that could capture data off-line in the field and then retransmit the information, when connected on-line, to the main database. In order to have a universal tool, we developed the app using an intermediate language that could be "translated" to Android and other operating systems such as iOS.

The progress with the web-based application was slow and, hence, a manual protocol was developed and technicians would visit each farmer three times in the cropping cycle (sowing, middle, end of harvest). The local technicians would record the information on the PF and would then load them onto their personal computers for uploading to the web-platform. Local technicians from FENALCE were trained in RASTA and the PF. Each field was georeferenced and given a unique identification number so that all information from one field could be accessed across years.

Weather information linked to individual fields during specific time periods was required. The maize growers we worked with, typically neither have weather stations nor regular weather reports they can access. Since 2014, the National Institute of Hydrology, Meteorology, and Environmental Studies (IDEAM) releases information from its meteorology stations on request (Young and Verhulst, 2017). IDEAM provided the full dataset of weather information from six weather stations in the study region. The daily weather station information was then linked to each of the individual fields. Details of the process to link weather information to each individual field are described in supplementary information section (S1.3).

Weather data was assigned to the distinct stages of crop development: vegetative 0–40 Days After Emergence (DAE), flower initiation to cob formation (41–97 DAE), and ripening (98 DAE to harvest). These time windows were estimated based on experiments conducted in Cordoba as farmers had no record of the phenological stage dates.

The web-platform SIRIA (http://siria.fenalce.org/) was designed to provide for safe data storage and reporting, with automatic validations for the units, ranges and a pool of options for each question, in order to reduce digitalization errors. The web-platform was initially developed by CIAT in Java Open JDK version 7, the records were stored in a MySQL database coupled with an Apache Tomcat, and they were exported to Excel via specific queries using the PhpMyAdmin interface. Data can be retrieved from the database using a dedicated query. Data were manually checked for common errors such as outliers, decimal notation or text in numerical variables. The dataset was consolidated with all variables involved (management practices, weather, soil, and yield), but without personal information from the farmers such as telephone numbers, address or e-mail.

When the processed dataset had missing values, either because they simply were not entered or because they were detected as outliers from the filters, those variables with more than 30% missing data were

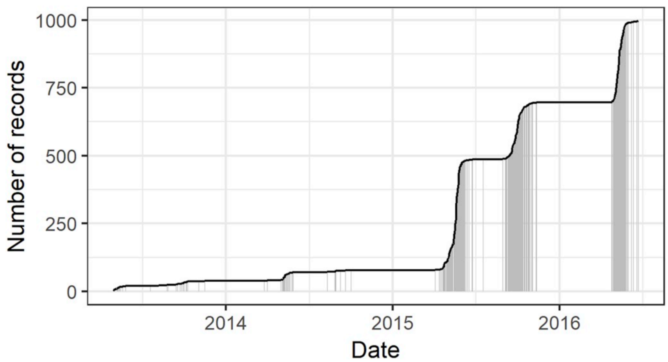

Fig. 1. Cumulative number of cropping events registered in the data base over time.

discarded. We noted few cases where the yield was extremely low due to crop failure: these extreme cases were removed. For the remaining variables, missing data were estimated with data imputation based on Random Forest (RF). This methodology estimates new values through a proximity matrix to other observations which are not missing (Breiman and Cutler, 2003).

2.2. Analysis and interpretation

Each and every time a farmer prepares a field, plants and manages a crop, he or she observes and experiments with a unique set of conditions (Cock et al., 2011). The analysis and interpretation are based on looking for associations between yield and the multiple variables in these individual experiments. Hence, there is no formal experimental design as is the case with, for example, randomized control trials which are particularly suited to evaluating the effects of a small number of variables. As we were attempting to discern the associations and interactions of multiple variables, the analysis of many individual events was essential. To analyze cropping events the dataset can be divided into subsets corresponding to homogeneous environmental conditions when there is considerable environmental variation (Jiménez et al., 2016,) but when environmental variation is limited the dataset can be analyzed as a whole (Jiménez et al., 2009). In the case of the department of Cordoba, agronomists suggested that variation in both climate and soil factors was limited in the periods when data was collected and the dataset was initially analyzed as a whole. We first performed a feature selection process to retain the most informative predictors from the set of variables integrated into the models (Table 1). A Random Forest algorithm (Breiman, 1999) was used to relate the yield to the various inputs and then determine which of those predictors are the most important drivers of variation in yield (Archer, 2008; Delerce, 2016; Strobl, 2008) . For more details see supplementary information (S1.4). As the results were analyzed and interpreted by agronomists, further analyses were made according to their insights in an iterative process.

In order to evaluate effectiveness of data-driven management recommendations, a low-cost participatory research experiment was established in farmers' fields. Six fields, which farmers were about to plant, were chosen. Each farmer reserved half a hectare to evaluate the data-driven management guidelines. FENALCE paid for any extra costs related to the recommended management practices. The practices for the data-driven subplots, were generated from the first round of analysis combined with advisory services recommendations.

2.3. Presentation of results to farmers

The results of the initial analysis and interpretation, developed and loaded on SIRIA, were discussed with the FENALCE agronomists. From these discussions the agronomists developed a series of guidelines for obtaining stable, high yields. A report called FENALCHECK (Fig. S6B), derived from SIRIA, was developed to indicate which factors were associated with high yields and to provide a checklist which farmers could use to rapidly see which improved practices they were not using. In addition, simple reports on the soil analysis were automatically produced by the SIRIA platform (Fig. S6A). Currently, farmers are not directly feeding their data into FENALCHECK. FENALCE technicians transferred the data from the PF to SIRIA which then automatically developed the FENALCHECK reports.

3. Results

3.1. Data collection and cleaning

The users of the app for data capture with an Android phone deemed it to be inefficient and difficult to use. The app was slow with inconsistent behavior and frequently froze. The targeted users did not adopt it. The CIAT-FENALCE team did not have the technical expertise to solve the problems. Later diagnosis attributed the difficulties that led to lack of adoption to the use of the intermediate language. The on-farm data was all collected using the specially developed PF. FENALCE personnel visited farmers three times during the growth cycle to fill in the PF. On the first visit to a field, farmers and the FENALCE personnel characterized the soil with the RASTA tool. The farmers trusted FENACLE and readily shared their data with FENALCE personnel. By mid-2015, with data collected over three years by FENALCE, about 400 cropping events were available for analysis in the department of Cordoba. (Fig. 1.).

Professional staff from FENALCE were trained in the use of SIRIA. The georeferenced information from the cropping events was digitalized and entered. This platform was used to manage information about farmers, farms, and fields within the farm. The original source-code for the web-platform was made available to FENALCE, which contracted specialized personnel to maintain, improve and administer SIRIA. At the time of writing SIRIA continues to be used by FENALCE. (S1.7). The data was cleaned with the automatic filters and validations reducing the number of errors. From the first 400 cropping events only 238 were suitable for analysis after cleaning the data and estimating missing values. Details of the processes for cleaning data are given in supplementary information (S1.5).

3.2. Analysis and interpretation

The observed farmers' yields ranged from total crop failure to more than 7 t ha^{-1}. The initial analysis of 238 cropping events and the RF model gave an R-squared of 46%. The main drivers of the yield included: the amount of phosphorus applied, the plant density 20 DAE, the run-off capacity of the field, total precipitation in the cob formation stage, and the seeding rate (sowing seeds number) (Fig. 2A.). The factors that can be easily managed by the farmer and that were most relevant for the model were: the amount of phosphorus applied, plant density 20 DAE, seeding rate, harvest method and cultivar.

The FENALCE extension agents saw how the information generated could be used to guide farmers in their management decisions. The FENALCE personnel combined the information generated by the analysis with previous knowledge to develop a comprehensive set of guidelines. In this process, the following guidelines were added: optimum range of seeding rates (distance between rows and plants within the row) and other practices needed to reach the optimum plant density at 20 DAE; practices to improve run-off; and more precise recommendations for fertilizer use.

From the initial analysis, FENALCE personnel developed five management practices to be evaluated in the on-farm trials. These were: (i) apply more than 0.015 t ha^{-1} of phosphorus, (ii) aim for a plant density (20 DAE) between 65,000 and 75,000 plants ha^{-1}, (iii) use mechanical harvest (iv) plant cultivar: Pioneer P3966, 30F35HRR or Pioneer

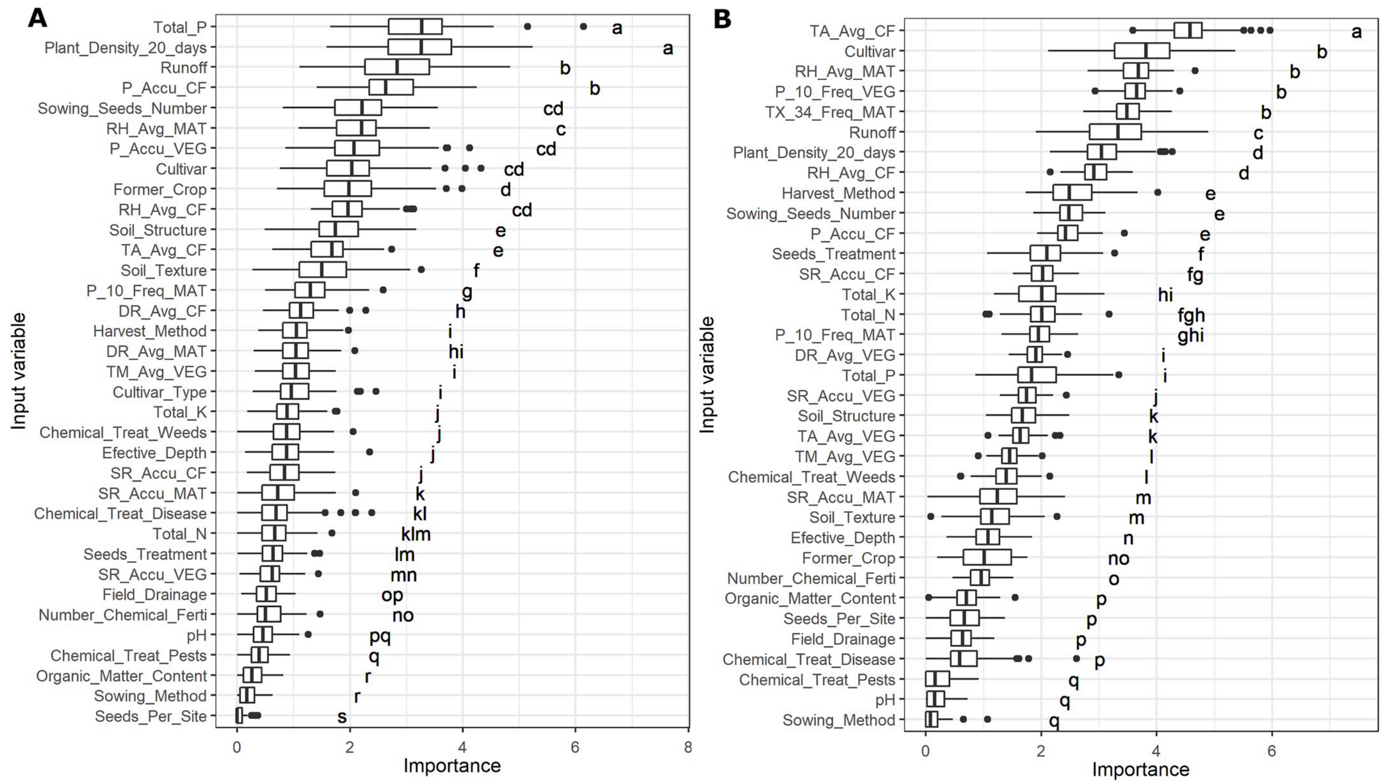

Fig. 2. A) Variable importance for the first model. B) Variable importance for the second model.

30F35H and (v) when intense rain is expected ensure that surface drainage canals and ditches are clean and in good condition. (Fig. 3.). FENALCE agents complemented these data-driven management practices with several mainstream recommendations: apply a minimum of 0.1 t ha^{-1} of nitrogen and use a series of practices including treated seeds to ensure an adequate plant population. Furthermore, the FENALCE personnel made several site-specific recommendations.

The on-farm trials faced several problems. One of the fields in the municipality of Cotorra flooded and the crop lodged as the crop was maturing. The farmer harvested manually, and did not separate data taken on the Data-Driven Technology (DDTech) plots from the controls. The plots from municipality el Tajo 1 also lodged, with heavier lodging in the demonstration plot. The plots were harvested manually and the DDTech plot yield, at 7.4 t ha^{-1}, was slightly greater than the control at 7.2 t ha^{-1}. In the remaining demonstration plots the yields ranged from 4.2 to 8.3 t ha^{-1} with the control plots consistently yielding less (Table 2).

The farmers from the first round of analysis were grouped into a Full group (F) who applied all the practices that were identified as advantageous by the analysis, None (N) for those who did not implement any of the practices, and Partial (P) were those that used some of the practices. Growers that used all five advantageous practices produced on average 2.5 t ha^{-1} more maize than those who did not use any (Fig. 4).

Up to the first analysis, and with the failure of the phone app, the data collection process was limited by the personnel dedicated to collecting data with the farmers. FENALCE realized that data-driven analysis of farmer's experiences could be used by its advisory services and hired extra personnel to intensify data collection.

At the end of the 2017 cropping season, the dataset had grown to more than 800 cropping events spanning four years (Fig. 1). A new round of analysis was run with the RF model on this expanded dataset. The R-squared increased to 66.5% with the weather-related predictors gaining importance in the explanation of the yield (Fig. 2B.) The management factors identified in the first round of analysis remained important but several factors that were previously of little significance were identified as important. The addition of more cropping events with more years of weather data did not increase the coefficient of variation of the weather variables (Table S2).

After identifying run-off and rainfall as relevant factors in the analysis, agronomists suggested that interactions between run-off and rainfall were likely. Consequently, the overall dataset was split into events with more and less than 400 mm of rain over the crop cycle and into sites with minimal external drainage (run-off) and those with moderate external drainage. The yield was greater with higher rainfall in the fields with moderate drainage (5.9 t ha^{-1}) than with minimal drainage (4.5 t ha^{-1}). The opposite was true with lower rainfall with higher yield (5 t ha^{-1}) with minimal drainage and (4 t ha^{-1}) with moderate drainage (Fig. 5.).

The field technicians modified their means of understanding and explaining variation in the field performance of maize as the program progressed. Initially, they used tacit knowledge, based on average performance in the area and general tendencies and trends over time, to explain variation in crop behavior. Later, they became aware of the importance of collecting data from individual plots and rigorous analysis to draw robust conclusions. The FENALCE technicians observed that several factors, which they believed to be important in determining yield, were not included in the data collected as they were difficult to measure. Nevertheless, the technicians were impressed by the ability of the analysis to explain yield variation.

3.3. Presentation of results to farmers

Originally, the program envisaged meetings of groups of farmers with an extension agent providing access to the SIRIA platform so that farmers would be able to discuss amongst themselves the best technology for their individual circumstances. This was not achieved. First, the structure and organization of the FENALCE advisory service was based on priority areas for each year. Hence, only a limited number of farmers from the study area were attended directly by the advisory service each year. The number of farmers reached each year directly by the sole FENALCE agronomist was 60. Second, FENALCE, was

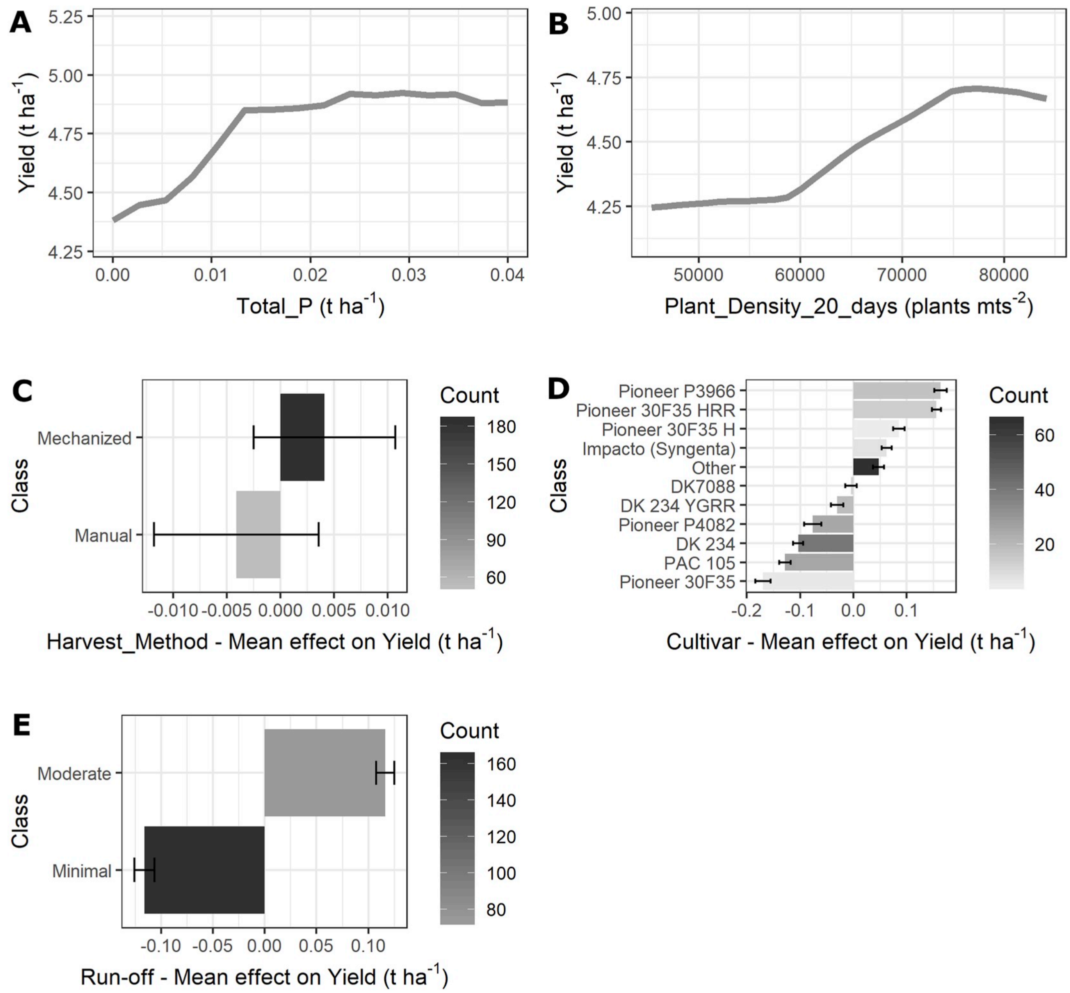

Fig. 3. Partial dependence plots of the most relevant predictors. A) Amount of phosphorus applied. B) Plant density 20 DAE. C) Harvest method. D) Cultivar. E) Run-off.

Table 2
Yield of the on farm-trial in Cordoba with farmer practices (control) and data driven recommendations (DDTech).

Site	Plot Yield (t ha $^{-1}$)		Comment
	DDTech	Control	
Chima	5.7	4.4	
Tajo 1	7.4	7.2	Lodged due to wind and heavy rain. Heavier lodging in the DDTech plot. Manual harvest.
Tajo 2	8.3	6.0	
San Antonio	6.8	6.0	
Carolina	4.2	3.7	
Cotorra			No data as farmer harvested fields by hand due to flooding and took no records.

concerned that farmers would derive mistaken conclusions from their direct access to the web-platform and this could lead to poor crop management or even failure. Third, the structure of the advisory service is not set up for group discussions.

The agronomist managing the advisory service indicated that he and those working with him had greater confidence in the trustworthiness of their advice, which was now backed up by data from commercial fields, and that this confidence was transmitted to farmers. Nevertheless, they observed that farmers paid little attention to the recommendations on improved drainage in the poorly drained areas.

4. Discussion and lessons learned

The United States Agency for International Development (USAID) appraised the digital agriculture program of CIAT in 2018 (Manfre and Laytham, 2018). We make considerable use of this independent evaluation in the following sections.

The approach of using information from multiple sources, including farmers' experiences, was demonstrated to be a useful means of providing extension agents and farmers with guidelines on how to improve production of rainfed maize in northern Colombia within the context of small-scale agriculture (Manfre and Laytham, 2018). The main features of developing such an approach are: coordination and collaboration between multiple institutions and individuals; data collection; data analysis and interpretation; and presentation of the findings to extension agents and farmers in an actionable format.

4.1. Institutions and collaboration

The program, from its inception, had political support from a high level in the main institutions involved in the program. The policy makers, principally in the MADR, provided a clear goal of strengthening the capacity to face weather and climate variability. Furthermore, the MADR designated the crop, the geographical area and the main institutions that were to develop the program. This clear definition of the objectives and major responsibilities, added to the stable funding expected from the MADR once the program strategy had been developed,

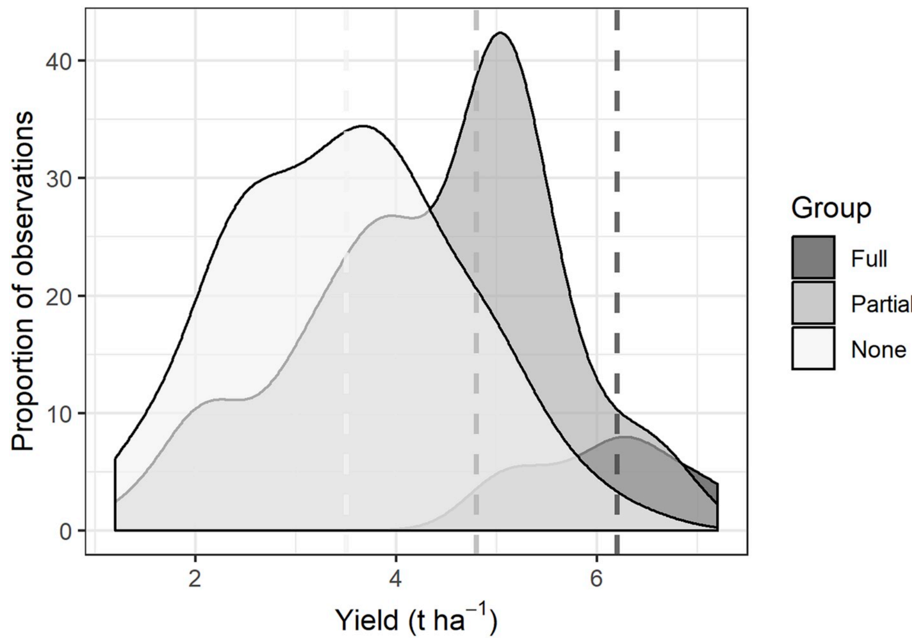

Fig. 4. Observed yield distribution for the farmers using all (Full group), part (Partial group) or none (None group) of the data-driven guidelines.

from our point of view, was key to success.

The MADR clearly defined what the objectives of the program should be, but they left FENALCE and CIAT with a relatively free hand to determine the strategy and approach. The flexibility allowed us to rapidly resolve difficulties when they arose. For example, the lack of a rigidly imposed work plan allowed us to react to the failure of the smartphone app for data collection with the development and use of the PF forms, which, whilst not ideal was pragmatic.

The two lead agencies (FENALCE and CIAT) executing the program had a clear alignment of interests. CIAT wished to further develop an incipient methodology for helping small-scale farmers to better manage their crops without going through the long-term traditional plot based research agenda, while FENALCE wished to provide their growers and extension agents with guidelines on how to better manage their crops to obtain higher and more stable yields. The common interest of the two agencies led to a harmonious working environment: when things went wrong, the attitude was not to determine who was to blame, but rather to find a solution. Thus, when the records kept by FENALCE were found to be insufficiently detailed to provide a meaningful analysis, CIAT, FENALCE and other agencies such as IDEAM all worked together to obtain the required data.

The Colombian government has stressed the country's commitment to open data, which favors the approach used in this program (Manfre and Laytham, 2018; Paul et al., 2018; Young and Verhulst, 2017). A striking example of this was the lack of weather data tied to individual fields at the start of the program. With its open data policies, IDEAM, the national meteorological agency, provided all the required data. In the program as a whole the collaboration from multiple agencies was critical (Esquivel et al., 2018; Manfre and Laytham, 2018; Ospina, 2018; Paul et al., 2018; Young and Verhulst, 2017). All collaborating agencies demonstrated a service mentality with all efforts directed to helping the maize growers to improve their crop management and hence food security and their livelihoods.

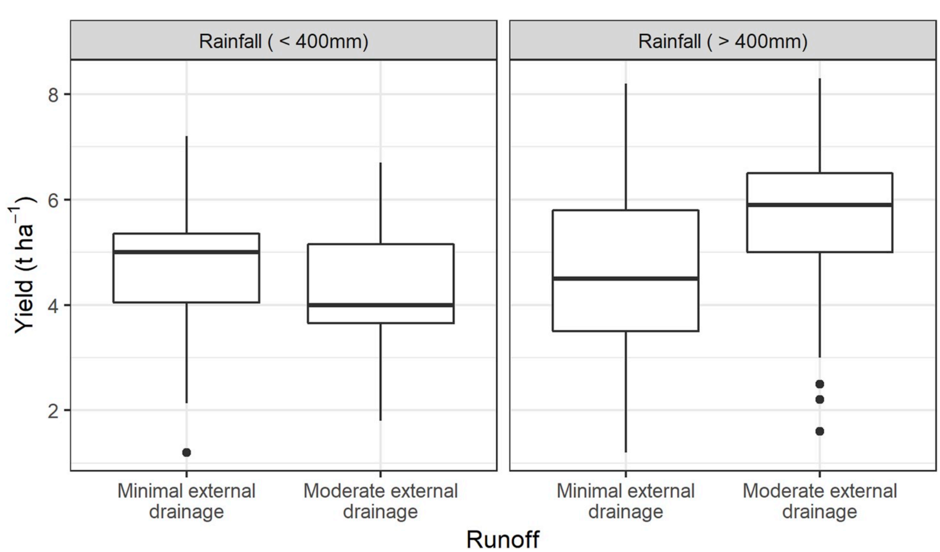

Fig. 5. Yield distribution across sites with minimal and moderate external drainage with rainfall above and below 400 mm during the cropping cycle.

4.2. Data collection

The lack of adoption of the data collection app hindered the progress of the program. The initial idea of farmers inputting their data directly on smart phones was not realized. In the future, we suggest, programs that wish to develop data collection apps for farmers should not develop apps from scratch, but should adapt existing apps to their needs as was done with Geo-Farmer (Eitzinger et al., 2019). The fall-back to filling in PFs enabled us to proceed, in a similar manner to past experiences (Jiménez, 2013), nevertheless to reduce the costs of data collection, avoid human errors in the collection, and to scale-up operations, new methodologies are required. Eventually, apps that farmers can use to input data, the Internet of Things (IoT) and automated collection of weather data and satellite imagery for crop development are some of the tools that can be used to improve the data collection process. Furthermore, low cost sensors are now available to monitor field conditions, drones can be used to monitor crop development, and combine harvesters with yield sensors can report yields of individual fields to centralized databases. These are just a few examples of the potential revolution in data collection in the coming years.

4.3. Analysis and interpretation

The yield on farmers' fields was extremely variable with a few farmers reaching more than 7 t ha^{-1}. The top yields are considered by extension agents to be good considering the resource base of soils, climate and infra-structure (e.g. drainage, irrigation) in the region. At the same time, the large number of farmers with low yields due to lack of good management of their crops indicates that there is a major opportunity to stabilize production at a higher level within the current restrictions of the resource base, and hence to contribute to food security.

The first round of analysis with 238 cropping events using machine learning algorithms identified topographical, management and weather variables associated with variation in yield. The initial analysis suggested that the yield could be increased by adopting five basic management practices. These findings led to two different opinions on how to proceed. One group wished to go ahead and recommend to farmers the practices identified and to combine them with their previous knowledge on good practices. This group felt there was no need to validate the practices as they had already been validated by the farmers that used them. The other, more cautious, group felt it would be unwise to go ahead with the recommendations without first testing their efficacy. In the end a compromise was reached in which the advisory service would use the information to make recommendations and trials would be established to verify the validity of the recommendations.

The on-farm trials were first envisaged as carefully controlled experiments, however, due to both budget restrictions and time limitations, only six on-farm trials were established. The minimum plot size was 0.5 ha, thus avoiding the problems of high yields associated with small experimental plots.

Despite the difficulties with water logging, the five comparisons between the recommended practices and farmers practices clearly showed a yield advantage when farmers followed the guidelines (Table 2). Farmers, unlike scientists, are normally more interested in the absolute yield level they obtain than the differences between treatments. The yields of the demonstration plots ranged from a low of 4.2 t ha^{-1} to 8.3 t ha^{-1}. These yields compare favorably with the estimated average yields for the zone of less than 4 t ha^{-1}.

A striking result of these trials was that three of them suffered from lodging or flooding. Lodging itself tend to be more severe when the soil is waterlogged. As noted, run-off was identified as one of the most important factors influencing yield (Fig. 3.) Furthermore, in the data cleaning exercises, fields that yielded very little or failed were eliminated from the analysis. Comments on the data collection PF indicated that the main reasons for elimination were excess water, strong winds and low rainfall. The elimination of events with low yields associated with excess water could have distorted the subsequent guidelines by underestimating the negative effects of low run-off. This then indicates that care is needed to avoid data cleaning that may eliminate data points as they may hold highly relevant data: other studies confirm the danger of eliminating data points that appear as outliers (Jiménez, 2018; Paul et al., 2018). At the same time the extension agents pointed out that the farmers did not pay attention to the guidelines related to improving drainage or not planting on sites with poor drainage. The trials highlighted the importance of these guidelines. Furthermore, we suggest that if the data elimination had not been practiced it is likely that the advisory service would have been more forceful in their appraisal of the perils of planting in poorly drained fields.

In the second round of analysis more variables were found that had a strong influence on yield and 66.5 % of the yield variation was explained by the variables used in the model. Weather variables increased their explanatory power in this second round. This would be expected as the addition of years may increase the overall variance of weather. However, this was not the cases (Table S2). This suggests that the explanatory power of the models used increases as more cropping events are analyzed, even when the overall variance of the individual variables used in the models does not increase. Hence, the greater the number of cropping events that can be analyzed the more robust guidelines that are generated. The crop management practices identified as important in the second round of analysis were consistent with the first round. Thus, generalized guidelines for management can be developed for use within this local context when the year-to-year weather pattern does not vary greatly.

The importance of weather-related predictors draws attention to the need for caution. In the period analyzed the weather patterns varied little from year to year as indicated by the similar coefficients of variance when extra years were added (Table S2). More data is needed from abnormal years to define which practices are optimum for distinct weather conditions. For example, we speculate that in a dry Niño year low run-off, which favors water conservation within a field, may be associated with higher rather than lower yields of maize as occurred in the years we analyzed. As more data becomes available, it may be necessary to separately analyze Niño, normal and Niña years with specific management for each condition. With the current forecasts of the Niño-ENSO farmers could plan their crop management according to the likely scenarios, rather than use standard, generalized packages and hence would be able to increase food security in those years with adverse weather conditions.

A major limitation to the approach of analyzing how management practices interact with weather conditions is the possibility, as was the case in our experience, that during the period of observation there is little annual variation in the weather with no strong Niño: this highlights the importance of continuous monitoring of the crop performance over time so as to ensure that the effects of climatic variation can be accounted for. At the same time, we point out that standard controlled trials do not consider the yearly variation in weather patterns, so farmers are frequently left with little information on how to vary their management practices according to variable weather patterns. The observational approach reported here ensures that farmers can have access to information that covers a wider range of variation than provided by the top-down approach.

Although there was little variation within the annual rainfall patterns, there was considerable variation in the weather conditions, particularly rainfall, of the individual fields observed both within and across years. The interaction between drainage and accumulated rainfall effects on yield demonstrates that the analysis can identify how management practices may vary according to the soil and terrain of the site and the weather conditions. In areas with lower rainfall, improving drainage to reduce run-off will likely reduce yield, whereas in areas

Table 3
Original and modified guidelines used by FENALCE after the analyzing farmers' results.

Original FENALCE Guidelines	Additional guidelines added after analyzing famer's results.
-Ensure that there is no hardpan and that soil compaction is not severe.	-When intense rain is expected ensure that surface drainage canals and ditches are clean and in good condition.
-Only plant when the effective soil depth is greater than 30 cm. -Plant in soils with pH > 5.5 y < 6.5. -Clean plots with herbicides before planting. -Work the soil to a maximum of 30 cm. -Inter-row spacing 0.75 m y 0.85 m, distance between plants in the row 0.17 m y 0.2 m.	
-Apply N,P,K,Mg,S. Note: no information on levels.	-Apply minimum of 0.015 t ha^{-1} of P at planting and 0.1 t ha^{-1} of N.
-Split N application at growth stage at planting (20%), V6 (40%) and V10 (40%).	-Minimum application of 0.1 t ha^{-1} of N at planting (50%) and V6 (50%).
-Aim for a plant population of 50–70,000 plants ha^{-1} 20 days after emergence.	-Aim for a plant population between 65.000 y 75.000. To achieve this: use mechanical control of weeds in the early growth period; use certified seed; calibrate seed drill; use treated seed; plant when soil is moist at a depth of 3–4 cm.
-Monitor the crop for weed, disease and pest control at least 6 times with the first monitoring within the first 20 days after planting. -Apply herbicides between 8 days before planting and 2 days after planting. -Control diseases 10 days before and at anthesis.	
-Harvesting: No specific recommendation.	-Whenever possible use a combine harvester.

with heavier rainfall improved drainage will increase yields. Thus, site specific guidelines, rather than blanket recommendations, can be developed.

In a somewhat similar exercise in sugarcane with larger farmers, observational data was used to determine environment (defined by climate and soils) by genotype interactions and hence to provide guidelines on the most suitable variety for specific locations (Cock et al., 2011; Isaacs et al., 2007). In addition, analysis of the observational data clearly indicated an interaction between fertilizer response, weather conditions, soil characteristics and topography with oil palm (Cock et al., 2016). Hence, we are optimistic that the proposed methodology will detect the interactions between weather conditions and management practices when there is a larger variation in weather patterns than that which was encountered in the years reported in this study.

Three basic pathways were used to move through the analysis and interpretation of the data towards subsequent establishment of guidelines for farmers. The guidelines were developed: (i) directly from the initial data analysis; (ii) through interpretation of the initial analysis by experts; and (iii) confirmation of expert recommendations from the data analysis. These three modes are illustrated by the following examples.

Previously FENALCE indicated the importance of balanced fertilizer applications, but did not provide guidelines on the levels of fertilizers to be applied. The initial data analysis indicated that to obtain good yields a minimum application of P was necessary (Table 3.). Hence, FENALCE now suggests a minimum dosage of 0.1 t ha^{-1} of P (Fig. 3.).

The initial analysis indicated a close negative relationship between slope and yield. This, initially, surprised agronomists. However, after discussions it appeared likely that the negative associations were due to accumulation of water in low spots and poor drainage. Further analysis using run-off as a variable, rather than slope, confirmed this hypothesis. Agronomists, then suggested that there would likely be an interaction between rainfall and run-off. This was then confirmed by further analysis. Thus, there is now an awareness of the importance of good drainage, especially when rainfall is expected to be heavy.

One of the well-established guidelines for farmers was to aim for a plant population of 65–75,000 plants ha^{-1} 20 DAE. The analysis confirmed both the importance of crop establishment and confirmed that the upper ranges suggested by FENALCE for plant population at 20 DAE were appropriate (Fig. 3B.). The analysis also indicated that failure to reach the required plant population was a major cause of low yields. Thus, a more comprehensive series of guidelines based on both previous knowledge and recommendations and the insights provided by the analysis were developed (Table 3).

4.4. Knowledge management

A major challenge with data analytics in agriculture is communication of the results to the end users in a way that facilitates their decision making. Initially our strategy was to use the web-platform, with the farmers in groups accompanied by extension agents discussing the results of the analyses and then drawing conclusions on how they could better manage their crops. This would then lead to a virtuous circle with farmers adopting new practices, analyzing their experiences and adopting those that were beneficial. For users to get excited about our approach and set this virtuous circle in motion, it was very important that the platform would offer valuable information to them in the form of user-friendly formats and reports. The FENALCHECK report (S1.6) was developed so that farmers could register which practices they used and then associate their crop performance with their use: it is based on the principles of CROPCHECK which has helped increase yields in several crops in Chile (Araya et al., 2010). Systems, like CropCheck, which continually monitor on-farm practices and output, can be used to evaluate new technology, much of which is likely to come from experimental stations and research institutes, as it is introduced by farmers on their fields.

FENALCHECK functioned in the sense that it helped agronomists and extension agents transfer successful management crop practices to the farmers. Nevertheless, we did not achieve the discussion and interpretation of the results by farmers groups with extension agents as originally planned. This was at least partially due to the fear that farmers might draw erroneous conclusions if they accessed the SIRIA platform directly without the presence of a facilitator. Furthermore, the FENALCE advisory system is not set up for group discussions with farmers. We suggest that future programs should set up mechanisms in their advisory services that allow farmers or farmers' groups to obtain information directly from the web-platform. These may include farmers' groups access to the web-platform with discussions in the presence of a facilitator, direct access to a user-friendly web portal or Interactive Voice Response (IVR) systems. The Colombian Sugar Cane Research Center, developed a portal through which growers can and do access information, obtained from both research plots and commercial fields, that helps them manage their crops better (Jiménez, 2013). Currently, FENALCE is developing an IVR systems through which farmers will be able to obtain advice based on the information available in SIRIA.

Despite some of the difficulties mentioned above, farmers expressed their confidence in the information they received because it came via FENALCE, which they trusted (Manfre and Laytham, 2018; Paul et al., 2018). Farmers generally trust their peers more than outside experts (Annan et al., 2016; Gray et al., 2018) and are more likely to believe in data generated on-farm than in experimental stations (Cock et al., 2011). As the guidelines come from analysis of observational information, farmers are likely to trust them and act upon them. The intermediaries provided cohesion and communication between farmers and experts and facilitated the adoption of new technologies and practices by farmers (Manfre and Laytham, 2018). The generally high yields obtained with the on-farm trials, even when using apparently traditional technology, strongly indicates that the knowledge gained by the advisory service on how to improve crop management was being transmitted to farmers effectively. This close relationship between farmers and advisory services and the confidence generated by sharing information and basing recommendations on the analysis of that shared information, whilst intangible, may be a key factor in increased, more stable yields.

Recent FAO and other agencies reports on digital agriculture in the developing world indicate that while traditional extension is a ponderous process relying on poorly paid extension workers who travel from farm to farm, digital technologies will speed up the transfer of information while improving its quality and relevance (Annan et al., 2016; FAO, 2019; Gray et al., 2018; Nayyar et al., 2018; Ramirez-Villegas et al., 2018; Salemink et al., 2017). We suggest that these digital technologies will make it possible to collect, analyze and return to farmers' information pertinent to crop management.

5. Conclusions

The data-driven agricultural program: (i) brought together information from multiple sources; (ii) analyzed and interpreted that data; and (iii) presented the knowledge generated to decision makers in a format that they could understand. The process of collecting data, interpreting the data and using the knowledge generated to assist farmers in making decisions requires collaboration between multiple agencies and individual farmers. For rainfed maize in Cordoba, in northern Colombia, the data-driven program demonstrated that much of the variation in yield was associated with variation in both weather patterns and management practices. Although the year-to-year variation in weather over the study period was relatively small, the data-driven program has the capacity to describe how the weather affects yield and how optimal management varies as the weather patterns change. From this analysis guidelines for farmers were developed following three pathways: (i) directly from the initial data analysis; (ii) through interpretation of the initial analysis by experts; and (iii) confirmation of expert recommendations from the data analysis. On-farm-trials confirmed that productivity could be increased by adopting good agricultural practices identified or confirmed by the program. The work of advisory services is facilitated by the extra knowledge they obtain from data-driven programs such as that described here. The extension agents have more confidence in making their recommendations and their discussions with farmers as the knowledge they share is backed up by data obtained from farmers' real-life experiences. We suggest that by combining the knowledge generated by data-driven agriculture with long-term weather forecasts associated with the El Niño phenomenon farmers can better manage their crops with varying weather patterns. This will contribute to both their prosperity and food security in the region. The basic principles established in this case study can be applied to crops produced by small-scale farmers in the developing world to increase their productivity and manage their crops under varying weather conditions with less risk of crop failure.

Declarations of interest

None.

Acknowledgements

This work was carried out under the ACLIMATE program (http://www.aclimatecolombia.org/http://odimpact.org/files/case-aclimate-colombia.pdf) with the financial support of the Colombian Ministry of Agriculture and Rural Development (MADR), the CGIAR Research Program on Climate Change, Agriculture and Food Security (CCAFS) under the project Towards a Digital Climate Smart Agriculture transformation in Latin America and the CGIAR Platform for Big Data in Agriculture under the community of practice Data-Driven Agronomy. Both CCAFS and the Platform for Big Data in Agriculture are carried out with support from CGIAR Trust Fund Donors and through bilateral funding agreements. For details please visit https://www.cgiar.org/funders/. The views expressed in this paper cannot be taken to reflect the official opinions of these organizations. We acknowledge the MADR of Colombia not only for the financial support but for its active participation in leading the initiative. The authors thank to the National Cereals and Legumes Federation (FENALCE) for their support, and for their contribution with data and insights for this study. Finally, we thank all the farmers who provided us with their data which was the basis of this study and the anonymous reviewers whose perspicacious comments have greatly improved the document.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.gfs.2019.08.004.

References

Annan, K., Conway, G., Dryden, S., 2016. African Farmers in the Digital Age: How Digital Solutions Can Enable Rural Development. Foreign Aff.
Araya, F., Acevedo, R., Cabello, M.C., Jaramillo, C., Gonzalez, I., Toro, M., 2010. CropCheck Chile:Sistema de Extension para el Sector AgroAlimentario. In: Fundación Chile en el Programa Cropcheck, Santiago de Chile, second ed. .
Archer, K.J., Kimes, R. V., 2008. Computational Statistics & Data Analysis 52, 2249–2260. https://doi.org/10.1016/j.csda.2007.08.015.
Bänziger, M., Setimela, P.S., Hodson, D., Vivek, B., 2006. Breeding for improved abiotic stress tolerance in maize adapted to southern Africa. Agric. Water Manag. 80, 212–224. https://doi.org/10.1016/j.agwat.2005.07.014.
Breiman, L., 1999. Random forest. Mach. Learn. 45, 1–35. https://doi.org/10.1023/A:1010933404324.
Breiman, L., Cutler, A., 2003. Manual on Setting up, Using, and Understanding Random Forests V4. 0. Univ. California, Dep. Stat.
CIAT-MADR, 2015. Logros y retos de la agricultura colombiana frente al cambio climático.
Cock, J., Kam, S.P., Cook, S., Donough, C., Lim, Y.L., Jines-Leon, A., Lim, C.H., Primananda, S., Yen, B.T., Mohanaraj, S.N., Samosir, Y.M.S., Oberthür, T., 2016. Learning from commercial crop performance: oil palm yield response to management under well-defined growing conditions. Agric. Syst. 149, 99–111. https://doi.org/10.1016/j.agsy.2016.09.002.
Cock, J., Oberthür, T., Isaacs, C., Läderach, P.R., Palma, A., Carbonell, J., Watts, G., Amaya, A., Collet, L., Lema, G., Anderson, E., Victoria, J., 2011. Crop management based on field observations: case studies in sugarcane and coffee. Agric. Syst. 104, 755–769. https://doi.org/10.1016/j.agsy.2011.07.001.
Cock, J.H., Álvarez, D.M., Estrada, M., 2002. RASTA Rapid Soil and Terrain Assessment: Guía práctica para la caracterización del suelo y del terreno. Cali, Colombia.
Cook, S., Cock, J., Oberthür, T., Fisher, M., 2013. On-farm experimentation. Better Crop 97, 17–20.
Delerce, S., Dorado, H., Grillon, A., Rebolledo, M.C., Prager, S.D., Patiño, V.H., Garcés Varón, G., Jiménez, D.,, 2016. Assessing Weather-Yield Relationships in Rice at Local Scale Using Data Mining Approaches. PLoS One 11. https://doi.org/10.1371/journal.pone.0161620.
Eitzinger, A., Cock, J., Atzmanstorfer, K., Binder, C.R., Läderach, P., Bonilla-Findji, O., Bartling, M., Mwongera, C., Zurita, L., Jarvis, A., 2019. GeoFarmer: a monitoring and feedback system for agricultural development projects. Comput. Electron. Agric. 158, 109–121. https://doi.org/10.1016/j.compag.2019.01.049.
Esquivel, A., Llanos-Herrera, L., Agudelo, D., Prager, S.D., Fernandes, K., Rojas, A., Valencia, J.J., Ramirez-Villegas, J., 2018. Predictability of seasonal precipitation across major crop growing areas in Colombia. Clim. Serv. https://doi.org/10.1016/j.cliser.2018.09.001. 0–1.
Evans, L.T., Fischer, R.A., 1999. Yield potential: its definition, measurement, and significance. Crop Sci. 39, 1544–1551.
FAO, 2019. Digital technologies in agriculture and rural areas - status report.
Gray, B., Babcock, L., Tobias, L., McCord, M., Herrera, A., Cadavid, R., 2018. Digital Farmer Profiles: Reimagining Smallholder Agriculture 91.
Herrero, M., Thornton, P.K., Power, B., Bogard, J.R., Remans, R., Gerber, J.S., Nelson, G., See, L., Waha, K., Watson, R.A., West, P.C., Samberg, L.H., Steeg, J. Van De, Stephenson, E., Wijk, M. Van, Havlík, P., 2017. Farming and the geography of nutrient production for human use : a transdisciplinary analysis. Lancet Planet Heal 1, e33–42. https://doi.org/10.1016/S2542-5196(17)30007-4.
Howland, F., Muñoz, L.A., Staiger-Rivas, S., Cock, J., Alvarez, S., 2015. Data sharing and

use of ICTs in agriculture : working with small farmer groups in Colombia. Knowl. Manag. Dev. J. 11, 44–63.
Hoyos, N., Escobar, J., Restrepo, J.C., Arango, A.M., Ortiz, J.C., 2013. Impact of the 2010–2011 La Niña phenomenon in Colombia, South America: the human toll of an extreme weather event. Appl. Geogr. 39, 16–25.
Ingram, J., 2008. Agronomist–farmer knowledge encounters: an analysis of knowledge exchange in the context of best management practices in England. Agric. Hum. Val. 25, 405–418.
Isaacs, C.H., Carbonell, J.A., Amaya, A., Torres, J.S., Victoria, J.I., Quintero, R., Palma, A.E., Cock, J.H., 2007. Site specific agriculture and productivity in the Colombian sugar industry. In: Proceedings of the 26th Congress International Society of Sugar Cane Technologists (ISSCT).
Jiménez, D., 2018. Everyone's Looking for a Data Unicorn. There's No Such Thing. Medium.
Jiménez, D., 2013. Site-specific Crop Production Based on Farmers' Production Experiences in Colombia. Case Studies on Andean Blackberry (Rubus Glaucus Benth) and Lulo (Solanum Quitoense Lam). Ghent, Belgium.
Jiménez, D., Cock, J., Satizábal, H.F., Barreto S, M.a., Pérez-Uribe, A., Jarvis, A., Van Damme, P., 2009. Analysis of Andean blackberry (Rubus glaucus) production models obtained by means of artificial neural networks exploiting information collected by small-scale growers in Colombia and publicly available meteorological data. Comput. Electron. Agric. 69, 198–208. https://doi.org/10.1016/j.compag.2009.08.008.
Jiménez, D., Dorado, H., Cock, J., Prager, S.D., Delerce, S., Grillon, A., Bejarano, M.A., Benavides, H., Jarvis, A., 2016. From observation to information: data-driven understanding of on farm yield variation. PLoS One 11. https://doi.org/10.1371/journal.pone.0150015.
Kiptot, E., Franzel, S., 2015. Farmer-to-farmer extension : opportunities for enhancing performance of volunteer farmer trainers in Kenya. Dev. Pract. 4524, 503–517. https://doi.org/10.1080/09614524.2015.1029438.
Lacy, J., 2011. Cropcheck: farmer benchmarking participatory model to improve productivity. Agric. Syst. 104, 562–571.
Landini, F., 2016. Problemas de la extensión rural en América Latina. Perfiles Latinoam. 24, 47–68. https://doi.org/10.18504/pl2447-005-2016.
Loboguerrero, A.M., Boshell, F., León, G., Martinez-Baron, D., Giraldo, D., Recaman Mejía, L., Díaz, E., Cock, J., 2018. Bridging the gap between climate science and farmers in Colombia. Clim. Risk Manag. https://doi.org/10.1016/j.crm.2018.08.001. 0–1.
Manfre, C., Laytham, W., 2018. USAID Digital Tools for Agriculture.Putting Data at the Service of Agriculture: CIAT Case Study. Washington, D.C., USA. https://www.usaid.gov/digitalag/ciat-case-study.
Mehrabi, Z., Jiménez, D., Jarvis, A., 2018. Democratize access to digital agronomy. Nature 555.
Nakano, Y., Tsusaka, T.W., Aida, T., Valerien, P., 2018. Is farmer-to-farmer extension effective? The impact of training on technology adoption and rice farming productivity in Tanzania. World Dev. 105, 336–351.
Nayyar, S., de Cleene, S., Dreier, L., 2018. Innovation with a Purpose: the role of technology innovation. World Econ. Forum 1–42.
O'Neil, C., 2016. Arms race: going to college?" In: Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. Crown Publishing Group., New York.
Ospina, A.V., 2018. Big Data for Resilience Storybook. Experiences Integrating Big Data into Resilience Programming. Unlocking Big Data's Potential to Strengthen Farmers' Resilience. International Institute for Sustainable Development, Winnipeg (Retrieved from The Internat).
Paul, A., Jolley, C., Anthony, A., 2018. Reflecting the Past, Shaping the Future. USAID. Washington, D.C., USA.
Ramirez-Villegas, J., Jiménez, D., Lougee, R., 2018. Data-driven farming proves fertile ground for O.R. OR/MS Today 20–23.
Ray, D.K., Gerber, J.S., Macdonald, G.K., West, P.C., 2015. Climate variation explains a third of global crop yield variability. Nat. Commun. 6, 1–9. https://doi.org/10.1038/ncomms6989.
Ricciardi, V., Ramankutty, N., Mehrabi, Z., Jarvis, L., Chookolingo, B., 2018. How much of the world's food do smallholders produce? Glob. Food Sec 17, 64–72. https://doi.org/10.1016/j.gfs.2018.05.002.
Rosegrant, M.W., Cai, X., Cline, S.A., Nakagawa, N., 2002. The Role of Rainfed Agriculture in the Future of Global Food Production.
Rosenheim, J.A., Gratton, C., 2017. Ecoinformatics (big data) for agricultural entomology: pitfalls, progress, and promise. Annu. Rev. Entomol. 62https://doi.org/10.1146/annurev-ento-031616-035444. annurev-ento-031616-035444.
Sagarin, R., Pauchard, A., 2010. Observational approaches in ecology open new ground in a changing world. Front. Ecol. Environ. 8, 379–386.
Salemink, K., Strijker, D., Bosworth, G., 2017. Rural development in the digital age. A systematic literature review on unequal ICT availability, ad.pdf 54 2012–2014.
Steinke, J., van Etten, J., Zelan, P.M., 2017. Erratum to: the accuracy of farmer-generated data in an agricultural citizen science methodology (Agronomy for Sustainable Development, (2017), 37, 4, (32), 10.1007/s13593-017-0441-y. Agron. Sustain. Dev. 37. https://doi.org/10.1007/s13593-017-0446-6.
Strobl, C., Boulesteix, A., Kneib, T., Augustin, T., Zeileis, A, 2008. Conditional variable importance for random forests. BMC Bioinformatics 11, 1–11. https://doi.org/10.1186/1471-2105-9-307.
Van Asten, P.J.A., Fermont, A.M., Taulya, G., 2011. Drought is a major yield loss factor for rainfed East African highland banana. Agric. Water Manag. 98, 541–552. https://doi.org/10.1016/j.agwat.2010.10.005.
Young, A., Verhulst, S., 2017. Aclimate Colombia: Open Data to Improve Agricultural Resiliency. Open Data's Impact. GovLab. http://odimpact.org/files/case-aclimate-colombia.pdf.