Makara

A tool for cotton farmers to evaluate risk to income

Ponce-Pacheco, Mario Alberto; Adla, Soham; Guntha, Ramesh; Aravindakshan, Aiswarya; Presannakumar, Maya; Tyagi, Ashray; Nagi, Anukool; Pastore, Prashant; Pande, Saket

DOI
10.1016/j.atech.2024.100759

Publication date
2025

Document Version
Final published version

Published in
Smart Agricultural Technology

Smart Agricultural Technology 10 (2025) 100759

Contents lists available at ScienceDirect

Smart Agricultural Technology

journal homepage: www.journals.elsevier.com/smart-agricultural-technology

Makara: A tool for cotton farmers to evaluate risk to income

Mario Alberto Ponce-Pacheco [a,*], Soham Adla [a], Ramesh Guntha [b], Aiswarya Aravindakshan [b], Maya Presannakumar [b], Ashray Tyagi [c], Anukool Nagi [c], Prashant Pastore [c], Saket Pande [a]

[a] Department of Water Management, Faculty of Civil Engineering and Geosciences, Delft University of Technology, Stevinweg 1, Delft, 2628, CN, Netherlands
[b] Center for Wireless Networks & Applications (WNA), Amrita Vishwa Vidyapeetham, Amritapuri Campus, Amritapuri, Clappana P. O.Kollam, 690525, Kerala, India
[c] Solidaridad Network Asia Limited, New Delhi, India

ARTICLE INFO

Keywords:
Data Integration
User-centric Design
Risk communication
Modeling
Irrigation Adoption

ABSTRACT

Smallholder farmers are critical to global food production and natural resource management. Due to increased competition for water resources and variability in rainfall due to climate change, chronic irrigation water scarcity is rising particularly in drought-prone regions. Improving the awareness of climatic risk to yields and incomes is critical to sustainable agricultural intensification. However, adopting a new technology represents a certain level of risk for the farmers, who invest time and economic resources in changing their practices. We have developed a mobile application, currently for cotton, that would allow farmers to actualize the risk of growing cotton. By implementing a sociohydrological dynamic model with a kernel principal component analysis structural error model, the software provides a risk forecast of the yield and profit the user can expect at the end of the season. The mobile app not only processes social and agricultural information provided by the user but also retrieves and continually updates climate datasets from the web, as well as market prices. The users can request the execution of the sociohydrological model to the servers from their own mobile devices. By following an agile methodology, the mobile app has been tested with ∼100 farmers in order to get feedback from real users; this brought the opportunity to redesign the functionality based on the correct understanding of information and, a fast and clear management of the tool and helping in the adoption of the technology. This was combined with existing knowledge around communicating risk by using multiple modes of communication - text, graphics, sound and video - all of which were implemented to reinforce the knowledge communicated and ensure sufficient redundancy. This turned out to be beneficial for farmers with low prior knowledge and higher acceptability of the mobile app by the users as evidenced through feedback rounds with them. This study exemplifies an approach to address the gap in communicating risks in agriculture using a user-friendly mobile application.

1. Introduction

The effectiveness of agricultural extension can potentially increase with the integration of technologies such as mobile phones [1]. For example, mobile phones can improve agricultural extension service delivery and potentially catalyze improvements in farm productivity and rural incomes [2,3]. There is a high diffusion of smartphones in the world, with one-third of the global population owning a smartphone, more than half the population of the world connected to the internet, and mobile subscriptions having reached around 7.76 billion in 2017 [4].

Studies have investigated the widespread usage of mobile phone applications (apps) and other software tools for agricultural purposes [4–7]. Eichler Inwood and Dale [8] have described software tools from the perspective of farm or farmer-relevant functions such as regulatory compliance, information management, agronomy (for profitability, reference information and sensed data), product tracking and emissions accounting. Another categorization can be based on where the mobile app intervenes within the stage of the agricultural value chain - (i) pre-harvest (inputs and knowledge); (ii) harvest and transportation; (iii) processing and storage, and (iv) distribution, packaging and handling of finished goods [6].

Agriculture comprises complex interconnected systems and their corresponding uncertainties [9]. Risk is inherent in agriculture, and risk

* Corresponding author.
E-mail address: m.a.poncepacheco@tudelft.nl (M.A. Ponce-Pacheco).

https://doi.org/10.1016/j.atech.2024.100759
Received 5 November 2024; Received in revised form 24 December 2024; Accepted 25 December 2024
Available online 27 December 2024

management is an integral part of farm management to balance profits and risks, reducing losses, and utilizing opportunities [10]. Agricultural risks can arise from diverse sources, such as changing (input and output) prices, agricultural and environmental policies, global trends in consumptions, global markets, and climate and biological variables [11]. While risk assessment and modelling can be essential in supporting decision-making at the farm-level, these require objective data which can be sparse, and subjective data can be flawed in this context [12]. A better understanding of risk and risk management can help not only farmers in making better agricultural decisions, but also policymakers in assessing different risk protection tools [13].

Hence, progress in mobile technology in agriculture needs user-friendly applications which communicate quality, trustworthy and timely advisories [3,4], while also incorporating the corresponding risks.

Despite the multitude of mobile phone applications in agriculture, more research is needed on specific applications to reduce risks related to agricultural activities. Any mobile app on precision agriculture or farm management (such as those reported in [5,14]) can potentially support the reduction of farm risk, but may not address it directly. Applications have been able to support farmers in crop choice based on weather and market conditions, potentially reducing risk [15].

A mobile phone application for farmers to visualize and communicate agricultural risks is a logical next step forward considering interconnected agricultural systems, the inherent risk in agriculture, and the growing potential for the use of agricultural mobile applications. This study describes the development of the Makara app, aimed at communicating agricultural risks and risk mitigation strategies, which has been tested with (primarily) cotton-growing farmers in a drought-prone region of Maharashtra (India). Specifically, the area of study corresponds to 4 districts: Amravati, Nagpur, Wardha and Yavatmal (Fig. 1). Readers are directed to the literature for more details on the region whose data was used for model calibration [16–18].

2. Methodology

2.1. System overview

2.1.1. Front End features

Makara's Front End (FE) interface allows farmers to input the precise geographic coordinates of their farms, ensuring that location-specific climate and market information are utilized for tailored recommendations. The user can input farm-specific information, including plot size, soil type, irrigation methods, and crops they plan to sow. These include cotton, maize, and soybean, and can account for mixed or intercropping patterns. Farmers can store comprehensive expenditure details for inputs associated with each crop, and the mobile app maintains detailed accounts for each plot across cropping seasons. These inputs are automatically integrated into the risk communications.

When communicating risk, the factors that were relevant while devising the (yield, income and profit) risk communication strategy were the framing, format, and mode of communication. The framing of the risk communication was a major decision variable. It is important to account for the framing effect on the perception of the risks, i.e. the same prospects are perceived as more attractive if they are framed in terms of success rather than failure [19]. Another decision variable was the formats to communicate risks - numeric, verbal and visual [20,21]. Numeric formats involve risk communication via numbers or statistics (such as probabilities or likelihoods). Verbal formats use words or phrases to convey the likelihood or severity of the risk (e.g., "low risk"). Visual formats represent risk information graphically using diagrams, graphs or charts. Additionally, risk communication via multimedia technology (such as mobile phones) can employ different modes of communication, e.g., text, graphics, sound and video [22]. Other features, such as the farm expenditure table for financial record-keeping and the crop calendar and journal for tracking farm management activities in alignment with recommended good practices, further facilitated the farmers' usage of the mobile app.

The design of risk communication in this study was shaped through an iterative process of stakeholder engagement with about 100 farmers from the study region. The farmer stakeholders provided feedback that aided consequent modifications of the risk communication strategy. Five distinct feedback campaigns were conducted between March 2023 and January 2024, employing engagement methods such as observational walk-throughs, focused group discussions (FGDs) and individual feedback sessions [23,24], an example of which is shown in Fig. 2. These sessions, which were conducted through both in-person and virtual means, gathered stakeholder opinions on variables such as the ease of using the mobile app and the trust in its outputs. This approach ensured that the risk communication design was continuously refined and adapted to meet the actual needs and preferences of the farmers.

2.1.2. Back End features

The smallholder sociohydrological model (SHM) introduced by Pande and Savenije [25] was the basis of the model at the Back End (BE). The model is a dynamic system model that represents farm scale dynamics through the interactions between five state variables. The five state variables are the farmer's capital, soil moisture of the farm,

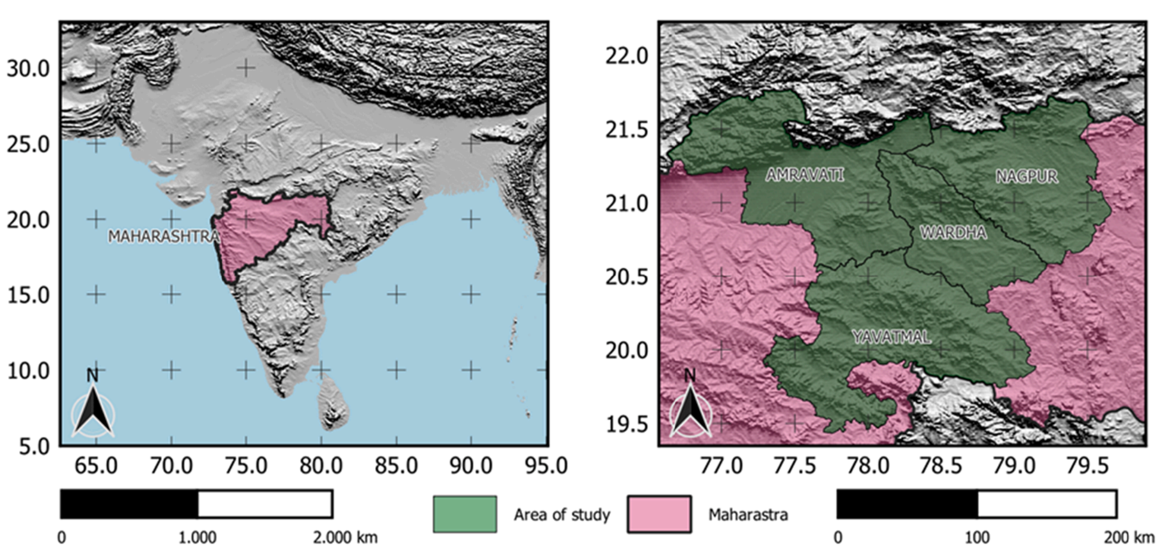

Fig. 1. Location of the area of study. left) Maharashtra state; right) 4 districts: Amravati, Nagpur, Wardha and Yavatmal.

Fig. 2. A photograph depicting the stakeholder engagement, taken during a focused group discussion with farmer stakeholders (March 2023). Photo courtesy: Solidaridad.

livestock if any, soil fertility that quantifies the nitrogen content of the soil, and fodder. These state variables are interlinked by various fluxes that then update the states of these variables. For example, soil moisture is updated after every time step (daily) by daily precipitation and evapotranspiration. The latter is determined by the crops grown by the farmer, which results in crop yields based on the evapotranspiration that is possible through the growing season(s). Farmers selling their produce in the market generate income flux, thereby increasing their capital just as expenditure on irrigation and other crop-related activities reduces their capital. The SHM also cuts on expenditure if the capital state variable of the farmers falls below zero, introducing a feedback loop of past poor yields on possibly low future yields due to cuts on farm inputs. Similarly, the farmer's fertilizer application linearly provides nitrogen inputs to the soil, updating the soil fertility state variable. For more details on the sociohydrological model, readers are referred to Pande and Savenije [25].

One critical sub-component of the model is how soil moisture and crop yield are interlinked. Pande and Savenije [25] used a single bucket model, which was enhanced by Djohan et al. [26], by using stress and biomass growth equations of the FAO AquaCrop model [27]. This study recoded the equations to be the same as those of the FAO AquaCrop model [27,28] and recoded the model to improve computational efficiencies. Further, Djohan et al. [26] also introduced a structural error model based on the regression of structural errors, i.e. discrepancies between yields observed in the 2019 household survey [16,18] and yields predicted by the SHM, with location-specific characteristics such as rainfall, soil types, farmer income, in a nonlinear kernel space [26]. The sum of yields predicted by the sociohydrological system dynamic model and the Kernel Principal Component Analysis (KPCA) machine learning-based structure error model (to be discussed in the methods section) results in a hybrid model that is then used for risk communication.

The hybrid model provides a yield prediction along with its uncertainty for each farm location based on spatial heterogeneities that it learns from the 2019 farm surveys. There are however additional uncertainties due to temporal variability in precipitation and temperature. Simulated yields are collated for years from the past with soil moisture conditions 'similar' to the soil moisture condition of a farm at the beginning of a simulation period. The overall uncertainty is then represented as the combination of spatial and temporal uncertainties. Each temporally 'similar' yield simulation also has associated spatial uncertainty as modelled by the hybrid yield prediction model.

The following section explains the sequence of communication between the FE and BE. It also outlines how the results are presented in the article.

2.1.3. Sequence of communication of the Makara App

Fig. 3 shows that the information management and model implementation in the mobile app are carried out on two different servers to separate the users' data from the model parameters and time series of climate data. The Front-End server processes and stores farmers' information in its database and the Back-End server scraps information from the web and runs the model.

Initially, the system administrator sets up the model parameters, such as the phenological parameters of the FAO AquaCrop model and socioeconomic factors. All this information is kept in the Back-End database and has to be manually updated by an administrator. In addition, the server retrieves climatic data from NASA-POWER [29] and crop prices from markets nearby [30] from the web automatically and continuously; this information is also stored in the BE database. In the Front-End database, users create accounts and can then generate Farm instances, which contain information such as location, soil type and best practices.

The execution of the model is requested by users from their devices directly to the FE server. Through the Django-REST [31], an Application Programming Interface (API) was developed to enable communication between the FE and BE servers. Thus, a service was set up in which by accessing specific URLs a request is submitted, executing concrete functions and returning an HTTP response. The users' devices communicate directly with the FE server, in which the information is processed and packed to be sent to the BE server using a Representational State Transfer (REST) approach, and the model is executed in the BE server. An inverse process is done to retrieve the model's output and deliver it to the user.

A Graphical User Interface (GUI) was developed and it is displayed on the users' mobile devices when they are using the mobile application. By using the mobile app, farmers can provide information about their farms and the practices they are engaged in. Some relevant information that the users provide are location, irrigation scheme, fertilization practices, crops grown and farm size; but also, some socioeconomic information can be retrieved such as family size, seeds and fertilizer costs, and costs of practices. This information is processed by the FE server and stored in the FE database.

The BE database contains information related to the climate, soil, and crop parameters. The soil dataset is directly populated from a geospatial source into a database that is geotagged. The climatic time series are obtained from the National Aeronautics and Space Administration (NASA) Langley Research Center (LaRC) Prediction of Worldwide Energy Resource (POWER) project funded through the NASA Earth Science/Applied Science Program [29], which is continuously updated by the system. This climatic information is processed to fill gaps. Since reference evapotranspiration (ET0) is not provided, it is estimated from the Top-Of-Atmosphere Shortwave Direct Normal Radiation and Temperature data [32]. Similarly, crop prices are constantly updated by web scraping from official sources due to the lack of a reliable API.

Since a part of the SHM involves the implementation of the AquaCrop model, the parameter values provided by the FAO's documentation [33] are used as initial values and calibrated for the region based on yields observed in the 2019 surveys.

The users access the system through authentication. In order to request the execution of the model in the BE from their mobile device, the users communicate directly with the FE server, in which the data needed for the model execution is pre-processed and packed. The information is then sent to the BE server in a JavaScript Object Notation (JSON) format through a POST, a standard HTTP method used to send data to a server for processing. A specific URL is defined in the REST environment to perform the model execution. Once this information is unpacked, it is used to link the parameters and time series in the BE database, by choosing the georeferenced information that is closest to the farm's location. The farmer's socioeconomic information is parsed together with the retrieved parameters and time series inputs into the hybrid model (composed of the calibrated SHM and the trained KPCA model). Finally, the probabilistic forecast is sent to the FE server, which stores it in the FE database and shows it in a graphical and audible way

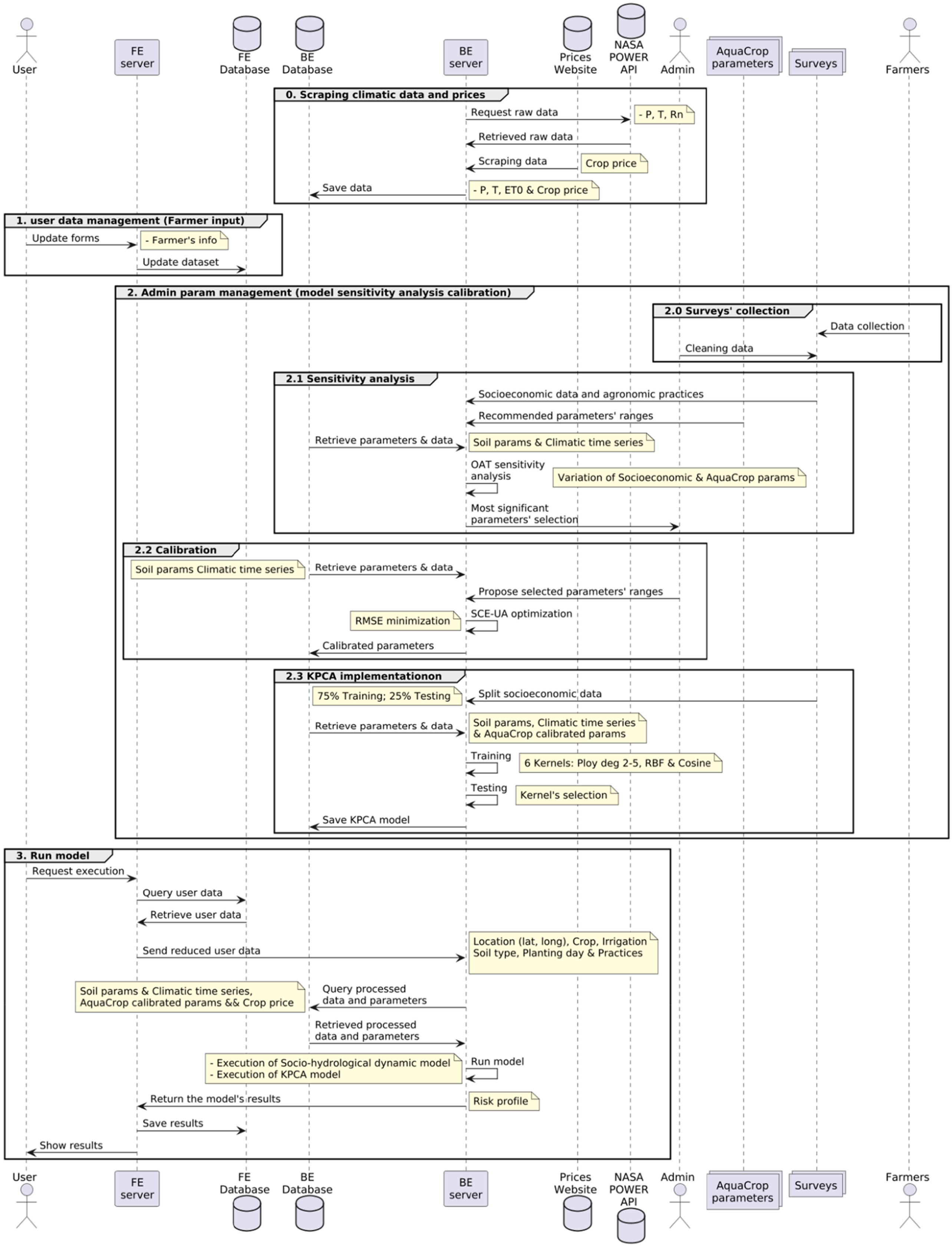

Fig. 3. The sequence diagram used for the Makara App. Web data is scraped and prepared, followed by user data management (including farmer inputs). Admin parameter management (model sensitivity analysis calibration) is performed in the next step, and finally, the model is run (simulation).

to the user.

2.2. Model input and management

2.2.1. Data scraping

2.2.1.1. Climatological data. The model uses climatological extracts of data extracted from the NASA-POWER Project [29] (see Table 1). The model uses raw precipitation and temperature data and processes evapotranspiration through the top of the atmosphere shortwave radiation [32] by using the calibrated Hargreaves and Samani equation [34, 35]. The temporal resolution of the model is daily, and the spatial resolution is 0.5° x 0.5°. There are time lags of 2 days for temperature and precipitation, and 180 days for solar radiation, which are filled by the respective historical averages corresponding to the missing calendar days.

2.2.1.2. Soil Data. The model uses soil properties such as depth, wilting point (WP), full capacity (FC), porosity and saturation coefficient (Ksat), Readily Evaporable Water (REW) and Total Evaporable Water (TEW) as parameters. This information is initially extracted and processed from the Indian soil dataset provided by the National Information System for Climate and Environment Studies (NICES) program of the Soil and Land Resources Assessment Division, Indian Space Research Organisation [36]. The used dataset has a resolution of 5 km x 5 km.

For soil depth, the original dataset provides a fraction of a grid that is in the ranges specified by bounds 0, 25, 50, 75, 100, 150, 200 cm and below. A representative value for every grid is obtained by applying the weighted average of the ranges (weighted by the fractions). To infer other parameters, the original dataset provides a fraction of the grid's area of a specific texture type and corresponding soil properties are then extracted.

Fig. 4 shows the spatial resolutions of precipitation (and other climate variables obtained from POWER dataset) and soil data. The grid nearest to the user's specified farm location is then used to obtain farm-specific data.

2.2.2. User input

Users may specify detailed inputs about their respective land, soil, and water-related characteristics. Land details include the location (retrieved from either their GPS location or via a manual search) and the area of the land. Soil characteristics include the soil depth, soil textural class and soil health parameters. For soil characteristics, the NICES dataset maintained by the National Remote Sensing Centre (Indian Space Research Organization, Government of India) is used to assign initial values of soil depth and soil textural class (see previous section). Users may overrule the soil textural class parameter based on their local farm locations based on the two prominent textural categories in the region, black cotton soil or red soil, assigned to clay and sandy clay textures, respectively [37]. Further, they are optionally prompted to add details of their soil health parameters including major and minor nutrients. Users are able to record their irrigation water sources (from choices including borewells, canals, pipelines, etc.) and technologies (including flood, drip and sprinkler). They are then prompted to add cropping choices and characteristics. Crop choices include cotton, maize and soybean based on local cropping choices, which were operationalized using AquaCrop parameters. There is an option of inter- or multi-cropping as well. Finally, to initialize the model, users may also input socio-demographic inputs such as livestock, family members, capital, loans and interest rates.

Table 1
Variables extracted from NASA-POWER.

Variable	Variable acronym	Unit	Lag (days)
Precipitation Total Corrected	PRECTOTCORR	mm/day	2
Temperature at 2 Meters	T2M	C	2
Min Temperature at 2 Meters	T2M_MIN	C	2
Max Temperature at 2 Meters	T2M_MAX	C	2
Top-Of-Atmosphere Shortwave Direct Normal Radiation	TOA_SW_DNI	MJ/m^2/day	180

2.2.3. Data management

2.2.3.1. Assigning data to farms. Due to the resolution of the soil data set (5 km x 5 km) and the climatic time series (0.5° x 0.5°), a strategy is needed to assign data to a geographical point defined by a farm's coordinates. A simple approach of choosing a soil and climate grid that is closest to the farm location is followed. This provides a reliable output to the users within a short time of processing, by establishing appropriate relationships in the database and bringing in the possibility to reuse queries in similar requests. A class 'Land' is defined that is linked to the closest Precipitation, Temperature and Evapotranspiration classes through a many-to-one relationship (Fig. 5). This relationship means, for example, that a Land can only contain a time series of Precipitation, but several Lands can share the same Precipitation time series. Similarly, an instance of class Land is linked to soil properties and soil depth. The coordinates are used to assign Land to a district, an administrative division. This allows scraping of crop prices based on the locations of nearby markets. Every time a user requests an execution of the model, the BE server finds the closest instances of the dataset and checks if an instance Land with these characteristics already exists. If this is the case, an instance of class 'Farm' is created and linked to the Land; otherwise, a new instance of Land is created and later linked to a Farm. Two instances of farms that are located relatively close share the same climatic time series and soil parameters; however, the distinction is defined by the crop cultivated and the type of irrigation and agricultural practices farmers implement.

2.2.3.2. List comprehension. In order to implement faster execution of the model, the model was encoded using list comprehension instead of other data structures because it 1) is Python native, and the addition of 3rd party packages is minimized; 2) represents a quick execution compared to other loops implementation; 3) allows conceptualizing of the model as a continuous series of values rather than year-defined stages, enabling seasonal data managemet even for agricultural seasons that spans across of two calendar years; 4) allows the processing of information through logical filters.

2.3. BE Model calibration and KPCA structural error model

Model calibration was conducted to obtain a set of parameters that leads to model simulations closely replicating the observed yields. This was done in two stages: parameter sensitivity analysis and then calibrating the sensitive parameters.

2.3.1. BE Model sensitivity analysis

Model sensitivity analysis is a procedure which estimates the rate of change in the model output as a response to the change in model parameters [38]. Initially, model constants and parameters were differentiated based on the SHM conceptualization. The constants were excluded from the calibration process, resulting in 15 parameters to be calibrated. The parameters related to the crop growth model were phenological parameters including time to emergence ('t_cc0′), time of senescence ('t_ss'), time to maturity ('t_m'); canopy development parameters like the canopy growth coefficient ('cgc'), canopy decline coefficient ('cdc'), initial canopy cover ('cc0_'); farm management parameters like the planting density ('density'); and others such as the initial harvest index ('hi_0′). Socio-economic and demographic parameters included the size of the user's family ('family_size'), number of

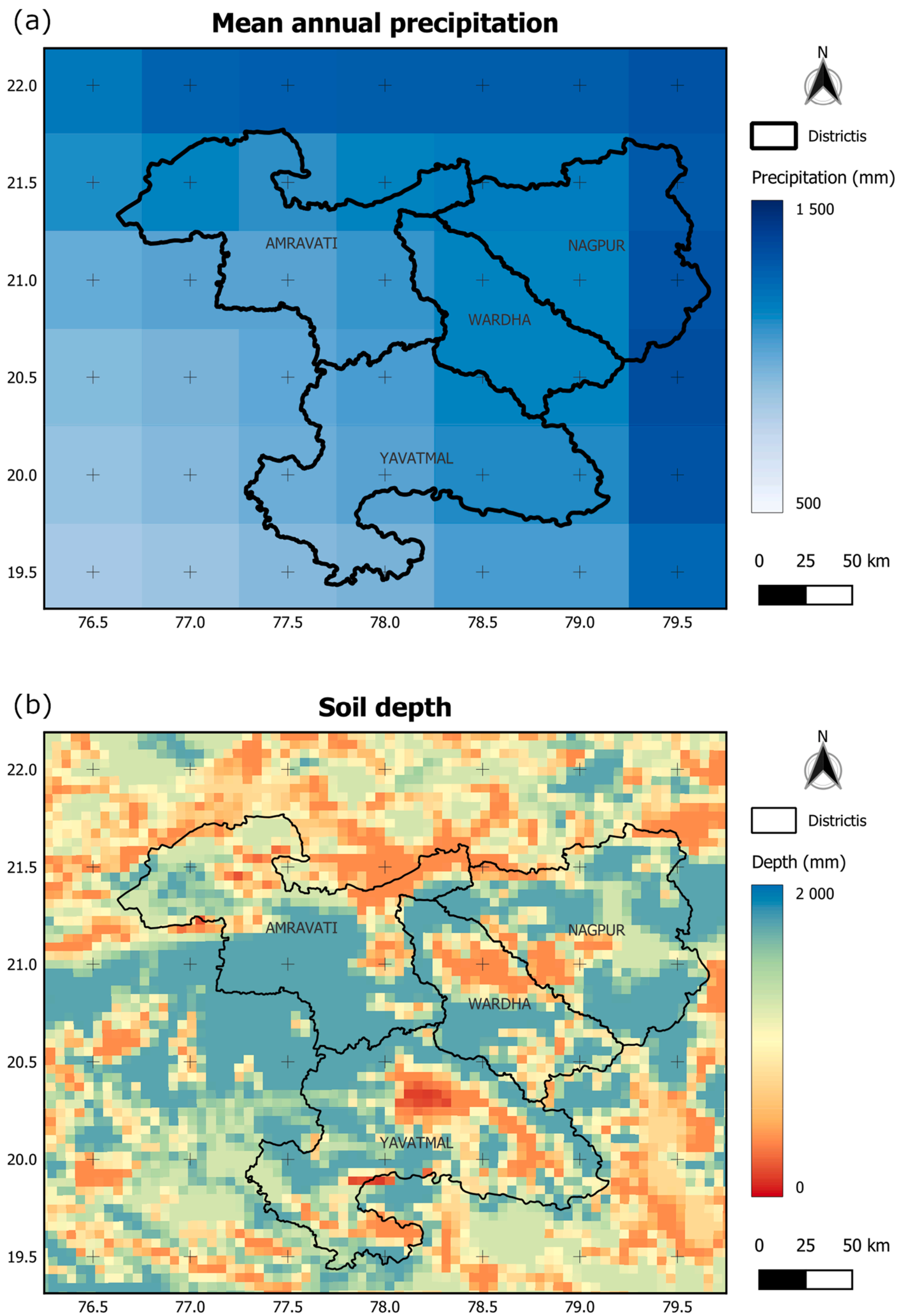

Fig. 4. Maps showing the gridded dataset of (a) mean annual precipitation (mm/year) and (b) soil depth data (mm). Weather data from NASA Power and soil data from NRSC had resolutions of 0.5° × 0.5° and 5 km × 5 km respectively.

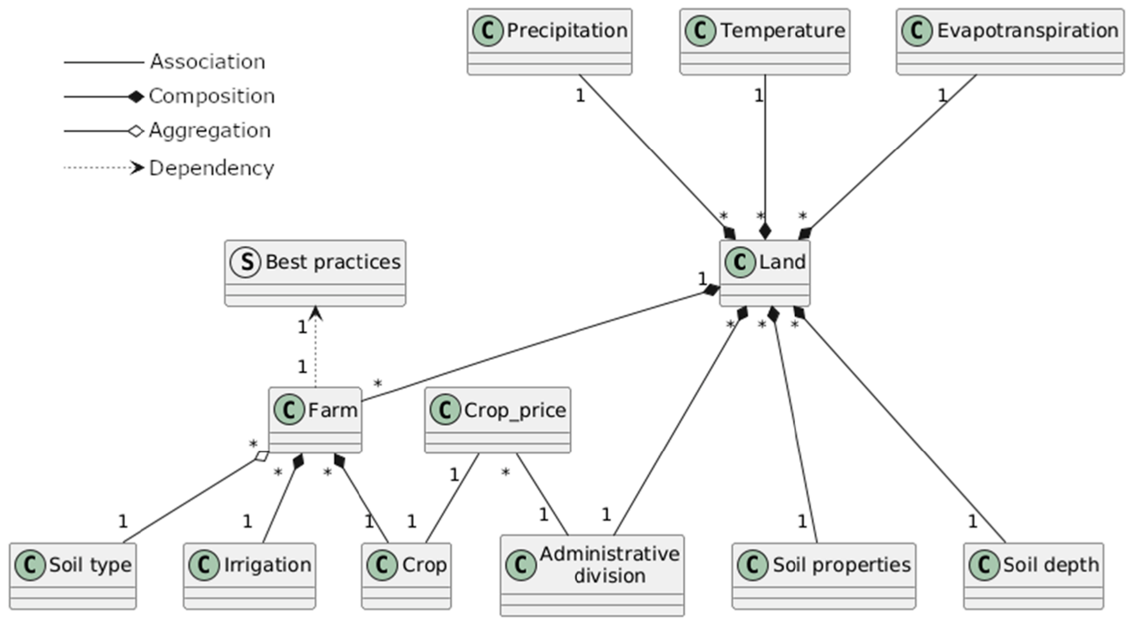

Fig. 5. The class diagram of data management of Makara App.

livestock ('livestock'), loan amounts ('loan_debt') and interest rates ('interest_rate'), and the prices of inputs such as seeds ('price_of_seeds') and fertilizers ('fert_price'), respectively and outputs such as the selling price of the crop ('price_of_crop').

SHM sensitivity was first determined globally, to estimate the change in the objective function as a function of changing each parameter, while all the other parameters were also changing [39]. Parameters were varied using the Latin Hypercube (LH) sampling technique [40] for the entire project area over a time series of 10 years/seasons of data. Each parameter was sampled between its minimum and maximum values (based on secondary literature, assuming a uniform probability distribution between the extreme values). The sensitivity index was the variation of the output yields (averaged across 10 years) corresponding to each parameter as a fraction of the global variation (considering all parameters). After the sensitivity indices of all parameters were computed, parameters were sorted in decreasing order of sensitivity, and then the parameters that explained ~98% of the total variance were selected for the next step of model calibration.

2.3.2. BE Model calibration

The Shuffled Complex Evolution - University of Arizona (SCE-UA) algorithm was used to optimize the sensitive model parameters identified during sensitivity analysis [41]. The SCE-UA combines the simplex procedure [42] with a controlled random search [43], competitive evolution and a complex shuffling concept [41]. This algorithm was selected due to its superior performance compared to other global and local search algorithms used to calibrate models [44]. The parameters identified as sensitive were optimized to minimize the objective function of the Root Mean Squared Error (RMSE, [33]), as defined in Equation 1:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \hat{x}_i)^2}{N}} \tag{1}$$

Where N is the total number of observations; $x_i$is the i^{th} observed value of x and $\hat{x}_i$ is the i^{th} predicted value of the same variable.

Table 2 provides the parameters of the search algorithm. These include parameters such as maximum number of function evaluations, number of complexes and other parameters linked to the convergence criteria of the algorithm.

Table 2
Parameters of the Shuffled Complex Evolution - University of Arizona (SCE-UA) algorithm used to optimize the sensitive model parameters.

Parameter	Description	Value
maxn	Maximum number of function evaluations allowed during optimization	10000
kstop	Maximum number of evolution loops before convergence	30
pcento	The percentage change allowed in kstop loops before convergence	0.001
peps	Value of NORMALIZED GEOMETRIC RANGE needed for convergence	0.001
iniflag	Flag for initial parameter array (=1, included it in initial population; otherwise, not included)	0
ngs	Number of complexes (sub-populations)	5
iseed	Initial random seed	0

2.3.3. KPCA structural error model

Djohan et al. [26] provide a machine-learning algorithm to model the part of the observed yield that is not explained by the SHM of Pande and Savenije [25]. The previous sections provide the calibrated SHM. The difference between thus calibrated SHM simulated yields for 2019 and observed yields at farms specified in the survey are regressed in a

Table 3
Description of the variables obtained from the 2019 survey [18] and climate forcing used in KPCA to develop the structural error model.

Variables	Unit	Description
Family help	person (s)	Farm labor help from the family/children
Cotton area	ha	Total cotton area of the farmer
Seeds cost	INR/ha	Local cost of cotton seeds
Pesticide cost	INR/ha	Local cost of pesticide
Fertilizer cost	INR/kg	Local cost of fertilizer
Fertilizer amount	kg	Total fertilizer usage of the farmer
Soil depth	mm	Soil depth in the field
Latitude	degree	Latitudinal coordinate of the farmer
Longitude	degree	Longitudinal coordinate of the farmer
Precipitation	mm	Total precipitation in the 2018-19 planting season
ETc	mm	Total reference evapotranspiration in the 2018-19 planting season
Irrigation	mm	The predicted cotton yield per hectare using the SHM
Model yield	kg/ha	Total irrigation in the 2018-19 planting season

nonlinear space as a function of variables shown in Table 3. Six different kernels (Radial Basis Function, Polynomial degree 2 to 5 and cosine kernels) are used to transform these variables in nonlinear space, where linear regressions are performed on principal components that significantly explain the variance of observed structural error.

3. Results

3.1. Model sensitivity analysis and calibration

3.1.1. BE Model sensitivity analysis

Fig. 6 illustrates the results of the sensitivity analysis of the SHM operating at the BE.

The y-axis of Fig. 6 shows fractions of total variability in yield explained by parameters in the x-axis. This fraction also has its own spatial variability (due to climatological data variability). 95% of variability is explained by t_ss, t_cc0. Further, only six parameters of the SHM (listed in Table 4) are chosen as they explained ~98% of the corresponding variability and then calibrated using the SCE - UA optimization. The SHM-modeled yields for the 2018-19 season from this calibration are then subtracted from the observed yields from the farmer surveys and KPCA is used to model the structural deficiency. The sum of the SHM prediction and the KPCA model of structural errors, i.e. the hybrid model, is then used as the predicted yield. For more details on the intermediate results, readers are referred to Djohan et al. [26].

3.1.2. Back-End Model

3.1.2.1. SHM calibration. Table 4 shows the best-fit values of the sensitive parameters after SCE-UA optimization. Descriptions are derived from the FAO AquaCrop manua [33]. References for default values are also provided within the table. Some phenological parameters have been transformed to ensure consistency in units. For the insensitive parameters, default values are used [33].

3.1.2.2. KPCA structural error model. Using the parameters above, the SHM simulates the yields of households reported in the 2019 survey [16, 18] and residuals obtained by subtracting it from the observed yields. These residuals were then used as dependent variables and regressed with nonlinear transformations of household-specific information of the variable reported in Table 4 that were found to be the significant predictors of the residuals [26]. Table 5 shows the performance of these regressions with various kernels on the training (75% of households), testing (25% of households in the survey) and overall data sets. The best-performing kernel RBF was chosen based on its performance on the test data set. RBF kernel has the lowest mean absolute error (MAE) of 371 kg/ha and the highest Nash Sutcliffe (NS) coefficient value of 0.183 compared to the other kernels on the test data. Polynomial degree 2 and cosine kernels are also close in performance with MAE and NS of 392 kg/ha and 388 kg/ha and 0.115 and 0.149 respectively and can also be used as kernels.

Using the RBF kernel, Fig. 7 (a) below shows the scatter of predicted residuals (i.e. prediction of structural errors) against the observed yield in the survey. The slope of the best-fit line (μ) is close to 1, which indicates low bias in the predictions. The r^2 shows that the RBF-based KPCA explains nearly 30% of the variance in observed residuals. Further, Fig. 7 (b) shows the residuals of these predictions, over which a

Table 4
Best-fit values of sensitive parameters after SCE-UA optimization.

Sensitive parameters	Definition of parameter	Units	Default value (reference)	Value after SCE-UA optimization
t_ss	Time between sowing and start of senescence (growing degree days)	°C	1400 [45]	1799
t_cc0	Time from sowing to emergence (growing degree days)	°C	45 [33]	78
cgc	Canopy growth coefficient (fraction ground cover increase per growing degree day during the canopy development phase)	°C-d^{-1}	0.007 [33]	0.008
density	Number of plants per hectare	plants/ hectare	105,000 [33]	108,632
hi_0	Reference harvest index (weight of lint and seed cotton divided by total dry biomass)	%	32.5 [33]	39.8
cdc	Canopy decline coefficient (fraction of ground cover decline per growing degree day after the start of senescence)	°C-d^{-1}	0.0025 [33]	0.0027

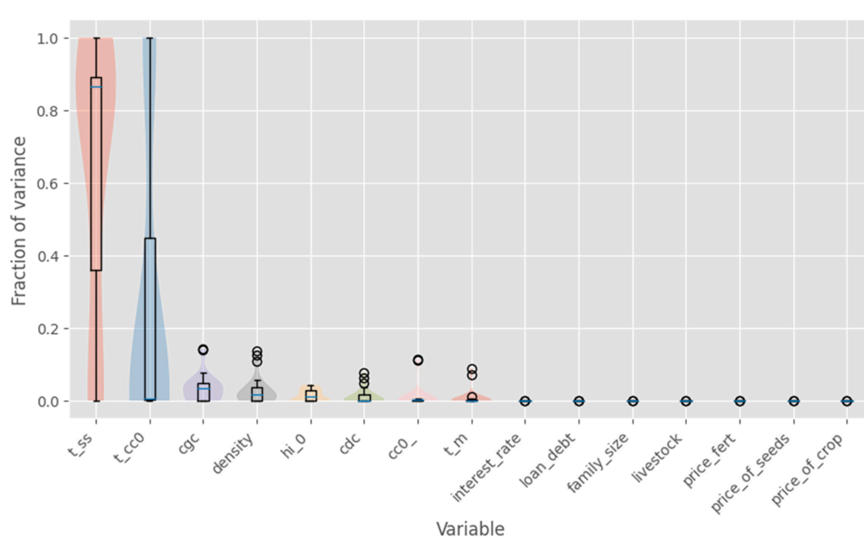

Fig. 6. Boxplots with density traces (using violin plots) depicting the global sensitivity of model parameters. The X-axis lists parameters and the Y-axis denotes the fraction of the total yield variability across the study area corresponding to the particular parameter. The full forms of the parameters are provided in Table 4.

Table 5
Results of the best-performing kernel in predicting observed SH model residuals (best structural error model)

	Training Data		Test Data		All Data	
	MAE [kg/ha]	NS [-]	MAE [kg/ha]	NS [-]	MAE [kg/ha]	NS [-]
RBF	291	0.336	371	0.183	311	0.296
Poly deg 2	322	0.221	392	0.115	340	0.196
Ploy deg 3	354	0.0372	416	0.00726	370	0.0352
Ploy deg 4	363	-0.00266	421	-0.0507	377	-0.00967
Poly deg 5	363	-0.00266	421	-0.0507	377	-0.00967
Cosine	321	0.212	388	0.149	338	0.2

Gaussian distribution is fitted to represent the distribution of a random variable, ϵ_r, such that the residuals are independent innovations of the random variables and drawn from the distribution. With the fitted mean, μ, of 40 kg/ha and standard deviation, σ, of 393 kg/ha, the distribution of the residuals has a small bias but high variance. This means that there are still several other unobserved variables not accounted for by the RBF-based KPCA model and as a result only ~30% of the observed variance is predicted by the model ($r^2 = 0.30$).

Fig. 8 shows the performance of the yield predictions by the hybrid model (i.e. the sum of the calibrated SHM and KPCA structural error model) when compared to observed yields. The r^2 shows that the hybrid model is able to explain 20% of the variation in observed data with uncertainty bounds (shown in yellow) covering almost 65% of the predicted data points. The slope of the best-fit line between the observed and the predicted is 0.80, showing the lower observed yields are overpredicted and higher values are overpredicted on average.

Fig. 9 depicts the spatial variability of mean yield predictions by the

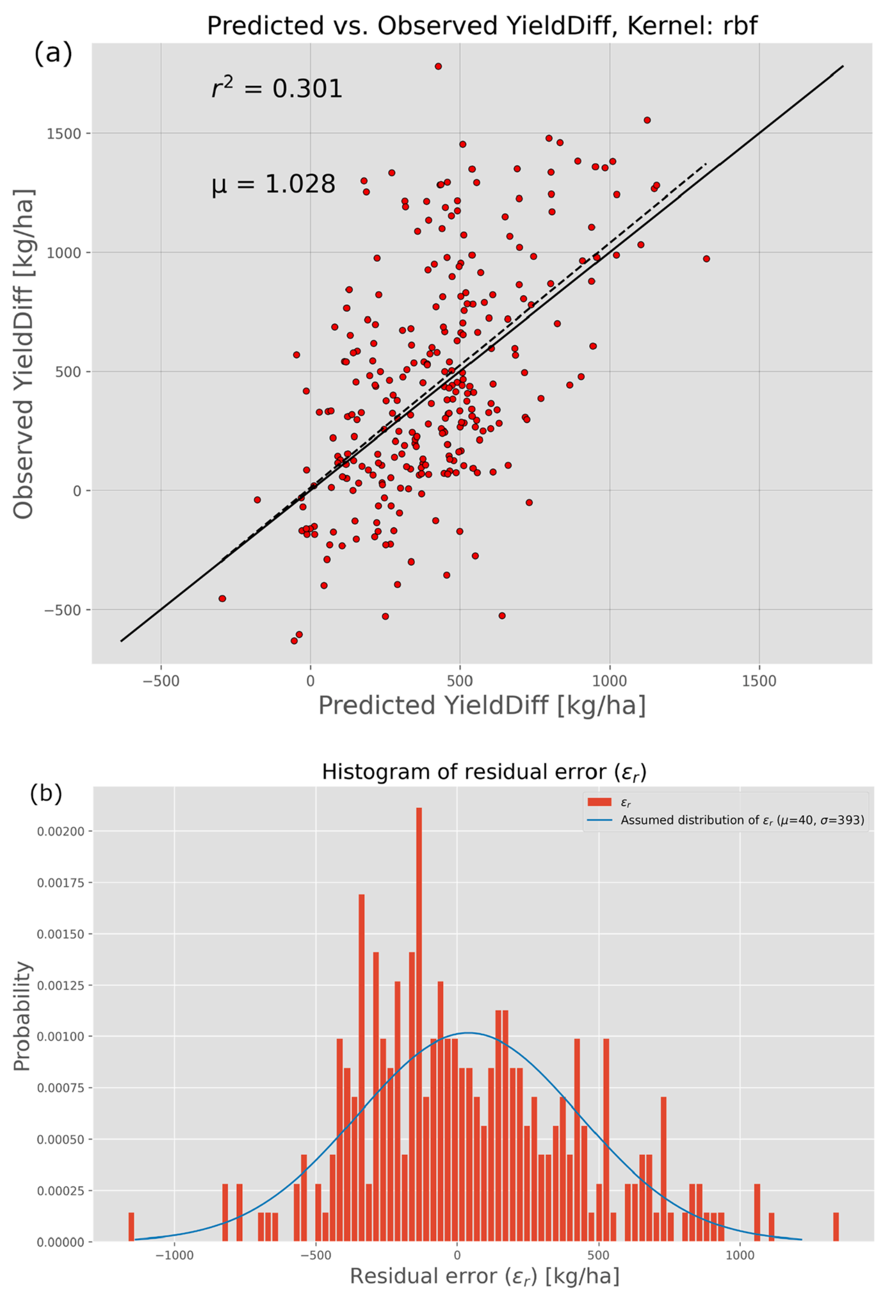

Fig. 7. (a): A scatter plot of the predicted and observed yield differences; (b) the histogram of the residual error (ϵ_r).

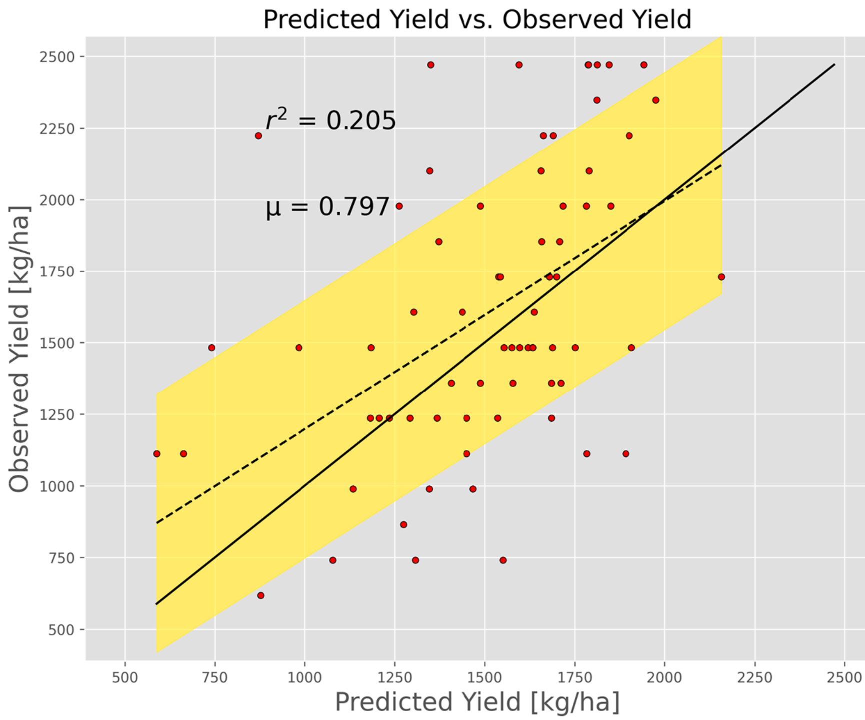

Fig. 8. A scatter plot between the observed and predicted yields using the hybrid model, along with the corresponding best-fit line.

hybrid model over the study region. It shows yield prediction under rainfed and under flood-irrigated conditions, in Figs. 9 (a) and (b), respectively. This shows that yield predictions under rainfed conditions are driven mostly by soil depths and rainfall patterns in space. The pattern smooths out, in particular, the effect of soil depth on yields, under flood irrigation conditions. This shows irrigation and smart advice towards good irrigation practices can make a big difference to sustainable farm incomes.

Risk prediction of Makara App is based on end-of-season yield prediction by the hybrid model and conditioned by local practices as inputted by the users (e.g. with respect to irrigation). Risk to yield, income and profit are represented by the 70% high probability region around the mean prediction as given by the uncertainty bound (yellow uncertainty interval in Fig 8). Also reported as part of the risk advisory are 95 percentile and 5 percentile yield values as the highest and lowest possible yields respectively. These percentile values are analytically obtained by adding mean prediction to corresponding quantiles of the Gaussian distribution fitted on the residuals in Fig. 7.

3.2. Makara v.1. features

The previous section presented the hybrid model, which is the predictor engine of the Makara App. Using the mobile app's screenshots (Figs. 10-13), this section elaborates on how inputs are received by the FE and the prediction is supplied back to the FE from the BE following the sequence diagram shown in Fig. 3.

3.2.1. Web data scraping and preparation

After creating an account, users start with no land. They have to create the land attributes and link them to other climate inputs (See Fig. 10). They start by specifying the latitude and longitude of their farm. This gives the geographic location for which relevant climate and

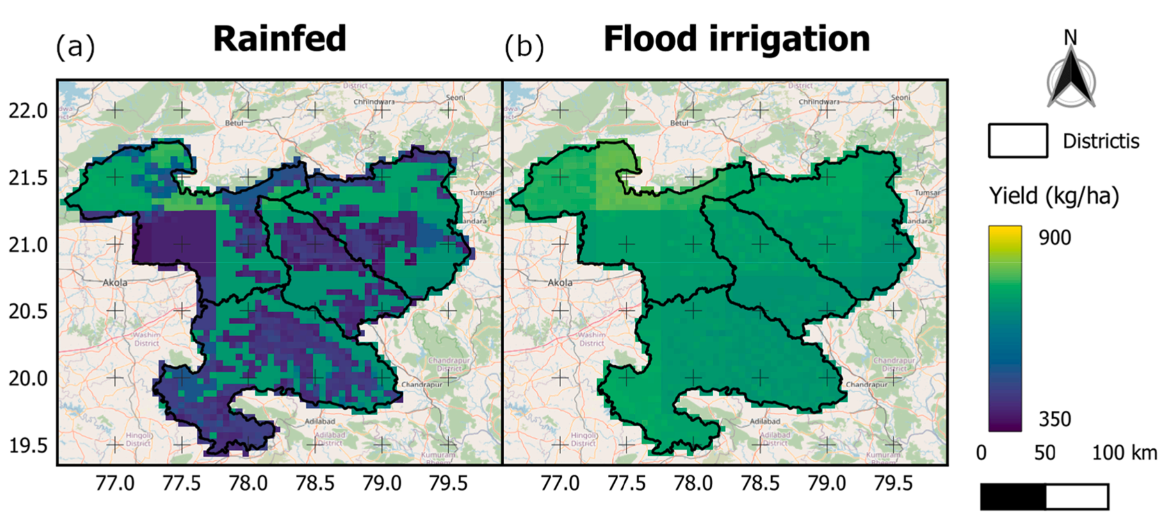

Fig. 9. Spatial variability in the yield predictions by the hybrid model over the study region under (a) rainfed, and (b) under flood-irrigated conditions.

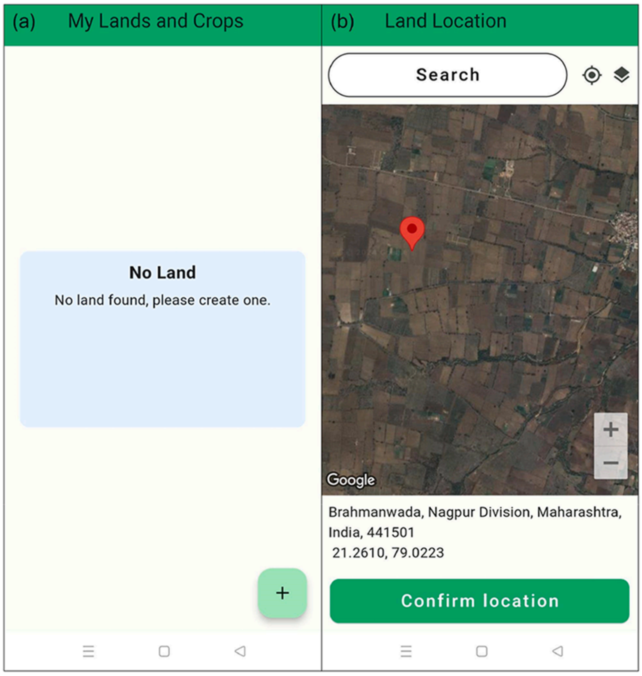

Fig. 10. Initial inputs for the Makara App: (a) adding a land and (b) specifying its location.

market data are harvested and populated in the BE database.

3.2.2. User data management (including farmer input)

They then further provide through the mobile app other farm-specific inputs such as the size of the land and the type of soil it has (Fig. 11a), soil fertility-related information (Figs. 11b and 11c) and water resources (Fig. 11d) that they can possibly use for irrigation. All this information is then used to populate the FE database of the class diagram shown in Fig. 5.

3.2.3. Admin parameter management (including model sensitivity analysis and calibration)

After having the Land database in the BE populated, additional crop-specific parameters are required in order to use the hybrid model for risk prediction (Section 2.3). This is achieved by the Makara App requesting farmer inputs on the crop growing season and crop types. They are then further requested to input whether there is multiple or intercropping and if so how are different crops arranged within. These are communicated back to the BE server, where crop-related parameters are further mined from an already calibrated database of the parameters of the hybrid model (results shown in Section 3.1.2).

3.2.4. Model run (simulation)

Once the location-specific parameters of the hybrid model have been assigned, the model simulates the yield, income and profit at the end of the growing season. The latter two are based on yield predictions, the price of cotton from the nearest market and agricultural expenses input by the farmers. This time, rather than farmers providing inputs to the BE, they receive risk advice based on model simulation from the BE. The GUI through which they receive advice was devised after interactions with 100 farmers (see methods section). The selected strategy delivers forecasts for crop yield, income, and profit, presented in a user-friendly audio-visual format.

A numerical format was selected, which communicated the ranges of yields, incomes and profit which had a 70% probability of occurrence based on the backend model calculations. This positive framing was preferred over framings in terms of failure, e.g. probability of falling below a certain value, to nudge farmers' perceptions towards opportunities instead of failure [46]. This aligned with the general recommendations for risk communication to preferably be exact, simple, concrete and relatable in terms of the information provided [21]. Moreover, multiple modes of communication were used - text, graphics, sound and video - all of which were aimed at reinforcing the knowledge communicated by each other. The automated voice would read out the content of the risk communication, while also establishing the broader context. This was in line with the claim that in audio-visual presentations, redundancy is beneficial for learners with low prior knowledge about communication [22].

4. Discussion

The risk communication relies on a hybrid model that is a mix of a sociohydrological system dynamic model and a KPCA-based structural error model. While the sociohydrological model considers the temporal dynamics of soil water and plant growth, the KPCA model learns the spatial patterns in the variance of observed yields unexplained by the sociohydrological dynamics as a function of various farm scale characteristics. These included cotton area, input costs and biophysical characteristics such as soil depth, temperature and precipitation. Such a hybrid model was able to explain 20% of the observed variance of yields at plot scale. These are low-medium accurate, especially compared to studies that have used machine learning methods in predicting yields with very high-resolution datasets (e.g. see [47,48]). However, this study demonstrates better yield predictive performance at the farm scale when compared to other smallholder yield studies using crop simulation models and input data with resolutions finer than that used in this study [49,50].

The difference between these and the present study is that the former uses very high to high spatial resolution time series data of biophysical variables such as greenness, temperature and canopy cover, e.g. by using SkySat, PlanetScope and Sentinel-2 at 2m to 20m resolution data at various points in time during a growing season [47,48]. Meanwhile, the present study uses a sociohydrological model to ingest lower-resolution temperature and rainfall time series data and the average of these in addition to socioeconomic data for the KPCA model (NASA POWER data at 0.5 degrees). The intention of the present model is to provide forward-looking yield predictions and corresponding risk advisory as a function of agricultural practices so that farmers can be motivated to take up good agricultural practices. This is not explicitly possible in high-resolution imagery-based machine learning yield predictors.

However, the differences also highlight a way forward in improving the accuracy of the present model by ingesting higher-resolution datasets than those that are currently being used. This can be done for example by assimilating Sentinel- 2A data to update biomass states as simulated by the sociohydrological model (e.g. see [51]) as well as by using higher resolution predictors in the KPCA structural error model. Additional data on socioeconomic variables such as indicators of traditional or cultural practices as predictors may also help improve the structural error model. Multiple machine learning algorithms could also be compared to investigate if there are other machine learning models for structural errors that improve the accuracy of the yield predictions (e.g. [47]).

5. Conclusions

This study presents a risk communication application, Makara v.1., that transforms insights generated by a complex sociohydrological model into a format that is easily interpreted by low-technology literacy farmers. The model considers both farm-specific climatic, and soil information as well as user-provided irrigation practices and other relevant information to predict yields with uncertainty bounds. The predictions were based on a hybrid model that is composed both of the SHM and as well a machine learning algorithm (RBF Kernel PCA) based

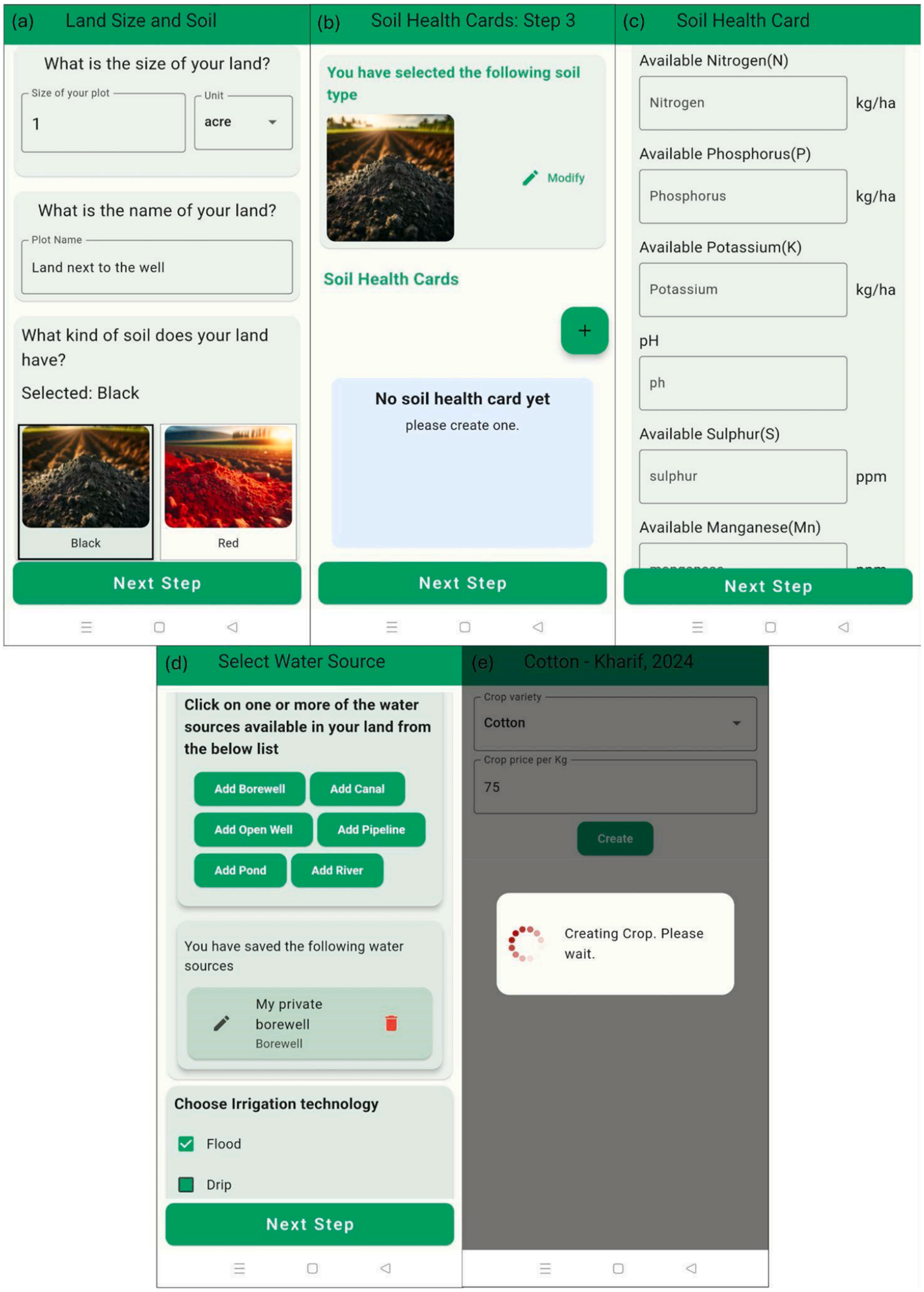

Fig. 11. Land-related inputs of Makara: details on (a) land size and soil, (b-c) soil fertility, and (d) water resources; (e) screenshot of the mobile app processing the data.

structural error model. The sensitive parameters of the SHM were calibrated using survey data collected from farmers in 2019. The difference between the farm-scale yield observations and the calibrated SHM-predicted yields was interpreted as a structural error. These structural errors were then regressed with additional socioeconomic and other location-specific variables in RBF kernel space to develop a KPCA structural error model. The sum of the SHM and KPCA-modelled yields then forms the hybrid model. The hybrid model also provides uncertainty bounds, on the basis of which a farmer's risk to yield, income and profit is communicated.

One unique contribution of the Makara App is the risk prediction engine at the BE that predicts yields and their uncertainty at the farm scale, which this article demonstrates. This forms the basis of a risk advisory that is informed both by farm-scale water human dynamics as well as machine learning algorithms. Yet another contribution of the mobile app is the user-friendliness to input farm scale information that contextualizes risk predictions better and the communication of the risk to low technology literacy farmers. The former is equally important, e.g., with respect to irrigation practices, which the study demonstrates has a significant effect on predicting yields and risks at farm scale. The article also reports the design methodology for a user-friendly GUI that required several rounds of feedback with about 100 farmers from the study area. This was combined with existing knowledge around communicating risk by using multiple modes of communication - text, graphics, sound and video - all of which were aimed at reinforcing the knowledge communicated by each other. This brought in sufficient

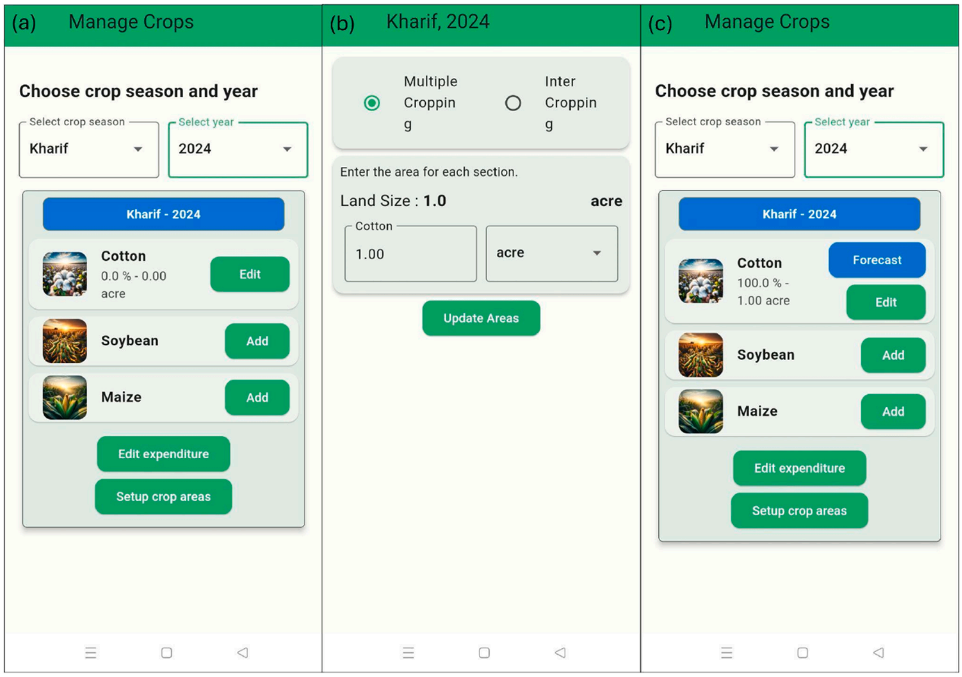

Fig. 12. Crop management in Makara - (a) adding crops, (b) specifying inter/multi-cropping and relative area of each crop, and (c) screen prior to forecast.

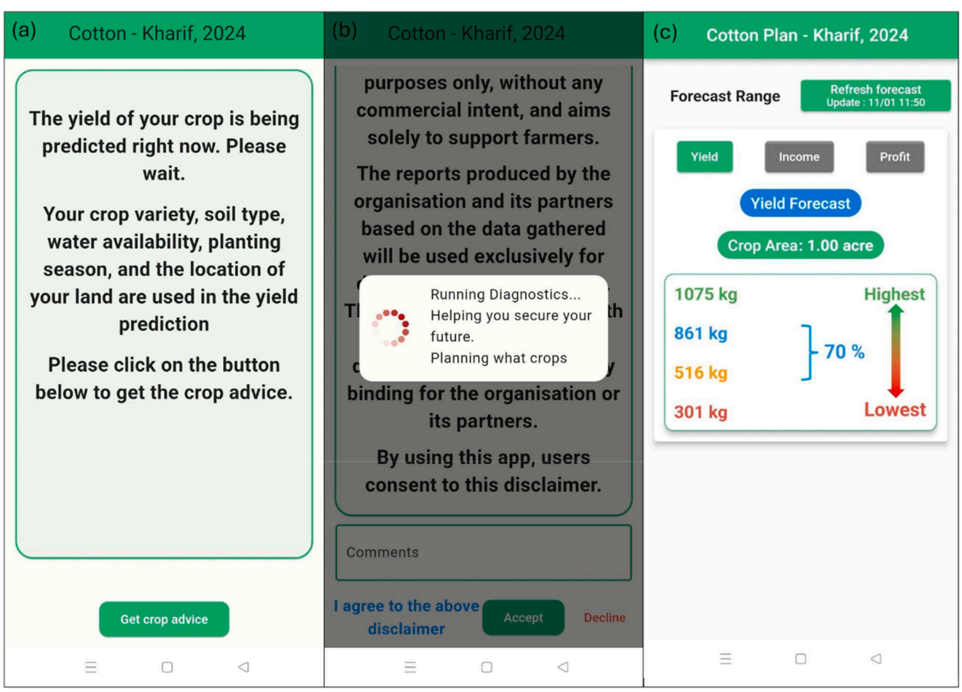

Fig. 13. The prediction screens of Makara, enabled in 3 languages with a voice-over - (a) contextualization of the risk assessment methodology, (b) disclaimer text and consent, and (c) the risk communication, which highlights a range of yield which has a 70% probability of being achieved.

redundancy that turned out to be beneficial for learners with low prior knowledge about the communication and for higher acceptability of the mobile app by the farmers as evidenced through feedback rounds with the farmers. The Makara App is currently undergoing field trials in a real operational environment with ~600 farmers in the study area, the feedback of which would verify the regional scalability of the mobile app to wider farmer populations and be presented in future research.

Ethics Statement

Not applicable: This manuscript does not include human or animal research.

CRediT authorship contribution statement

Mario Alberto Ponce-Pacheco: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Soham Adla:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Methodology, Conceptualization. **Ramesh Guntha:** Software, Methodology. **Aiswarya Aravindakshan:** Software, Methodology. **Maya Presannakumar:** Software, Methodology. **Ashray Tyagi:** Funding acquisition. **Anukool Nagi:** Funding acquisition. **Prashant Pastore:** Funding acquisition. **Saket Pande:** Writing – review & editing, Writing – original draft, Project administration, Methodology, Funding acquisition, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

[1] Z. Xu, A.E. Adeyemi, E. Catalan, S. Ma, A. Kogut, C. Guzman, A scoping review on technology applications in agricultural extension, PLOS ONE 18 (2023) e0292877, https://doi.org/10.1371/journal.pone.0292877.

[2] X. Fu, S. Akter, The Impact of Mobile Phone Technology on Agricultural Extension Services Delivery: Evidence from India, J. Dev. Stud. 52 (2016) 1561–1576, https://doi.org/10.1080/00220388.2016.1146700.

[3] Role of Mobile Phone Technology in Improving Small Farm Productivity, in: S. Mittal, G. Tripathi (Eds.), Role of Mobile Phone Technology in Improving Small Farm Productivity, Agric. Econ. Res. Rev. Agric. Econ. Res. Rev. (2009), https://doi.org/10.22004/ag.econ.57502.

[4] K.S. Kumar, C. Karthikeyan, Status of mobile agricultural apps in the global mobile ecosystem, Int. J. Educ. Dev. Using Inf. Commun. Technol. 15 (2019) 63–74.

[5] A.A. Aletdinova, Popular Mobile Applications for Crop Production, IOP Conf. Ser. Earth Environ. Sci. 666 (2021) 032036, https://doi.org/10.1088/1755-1315/666/3/032036.

[6] A. Anand, S. Raj, Agritech Startups: The Ray of Hope in Indian Agriculture (Discussion Paper No. 10), National Institute of Agricultural Extension Management (MANAGE), Hyderabad, 2019.

[7] A. Balkrishna, J. Sharma, H. Sharma, S. Mishra, S. Singh, S. Verma, V. Arya, Agricultural Mobile Apps used in India: Current Status and Gap Analysis, Agric. Sci. Dig. (2020).

[8] S.E. Eichler Inwood, V.H. Dale, State of apps targeting management for sustainability of agricultural landscapes. A review, Agron. Sustain. Dev. 39 (2019) 8, https://doi.org/10.1007/s13593-018-0549-8.

[9] C. Blackmore, N. Sriskandarajah, R. Ison, Developing learning systems for addressing uncertainty in farming, food and environment: what has changed in recent times? Int. J. Agric. Ext. 6 (2018) 03–15.

[10] R.B.M. Huirne, M.P.M. Meuwissen, J.B. Hardaker, J.R. Anderson, Risk and risk management in agriculture: an overview and empirical results, Int. J. Risk Assess. Manag. 1 (2000) 125–136, https://doi.org/10.1504/IJRAM.2000.001491.

[11] T. Bencová, A. Boháčiková, How to deal with global concept of Risk in Agriculture?, in: Comparative overview of the literature. SHS Web Conf. 92, 03001, 2021, https://doi.org/10.1051/shsconf/20219203001.

[12] V.A. Ogurtsov, M.P.A.M. Van Asseldonk, R.B.M. Huirne, Assessing and modelling catastrophic risk perceptions and attitudes in agriculture: a review, NJAS Wagening. J. Life Sci. 56 (2008) 39–58, https://doi.org/10.1016/S1573-5214(08)80016-4.

[13] Harwood, J.L., Heifner, R.G., Coble, K.H., Perry, J.E., Somwaru, A. (Eds.), 1999. Managing Risk in Farming: Concepts, Research, and Analysis, Agricultural Economic Report No. 774.

[14] J. Mendes, T.M. Pinho, F. Neves dos Santos, J.J. Sousa, E. Peres, J. Boaventura-Cunha, M. Cunha, R. Morais, Smartphone Applications Targeting Precision Agriculture Practices—A Systematic Review, Agronomy 10 (2020) 855, https://doi.org/10.3390/agronomy10060855.

[15] Ahmed, Faraaz, Ahmed, Faraz, Pasha, R., Prathap, V., Pasha, A., Kumari, D., 2018. Survey on Precision Farming using Mobile Applications.

[16] S. Adla, A. Šaponjić, A. Tyagi, A. Nagi, P. Pastore, S. Pande, Steering agricultural interventions towards sustained irrigation adoption by farmers: sociopsychological analysis of irrigation practices in Maharashtra, India, Hydrol. Sci. J. (2024), https://doi.org/10.1080/02626667.2024.2376709.

[17] N.R. Hatch, D. Daniel, S. Pande, Behavioral and socio-economic factors controlling irrigation adoption in Maharashtra, India. Hydrol. Sci. J. 67 (2022) 847–857.

[18] C. van Wirdum, N.R. Hatch, M.Y.A.M. Ali, P. Raghunathan, T. Willard, Multidisciplinary Project Cotton Water: Baseline study of designing sustainable instruments for smallholders in Maharashtra, India. Delft University of Technology, Delft, 2019.

[19] A. Tversky, D. Kahneman, The Framing of Decisions and the Psychology of Choice, Science 211 (1981) 453–458, https://doi.org/10.1126/science.7455683.

[20] I.M. Lipkus, Numeric, Verbal, and Visual Formats of Conveying Health Risks: Suggested Best Practices and Future Recommendations, Med. Decis. Making 27 (2007) 696–713, https://doi.org/10.1177/0272989X07307271.

[21] D. Timmermans, J. Oudhoff, Different Formats for the Communication of Risks: Verbal, Numerical, and Graphical Formats. Wiley Encyclopedia of Operations Research and Management Science, John Wiley & Sons, Ltd, 2011.

[22] Vetere, F., Howard, S., 2000. Prior knowledge and redundant multimedia, in: 2000 IEEE International Conference on Multimedia and Expo. ICME2000. Proceedings. Latest Advances in the Fast Changing World of Multimedia (Cat. No.00TH8532). Presented at the 2000 IEEE International Conference on Multimedia and Expo. ICME2000. Proceedings. Latest Advances in the Fast Changing World of Multimedia (Cat. No.00TH8532), pp. 605–608 vol. 2. 10.1109/ICME.2000.871436.

[23] B. Light, J. Burgess, S. Duguay, The walkthrough method: An approach to the study of apps, New Media Soc 20 (2018) 881–900, https://doi.org/10.1177/1461444816675438.

[24] H.R. Shibly, A. Abdullah, M.W. Murad, Research Framework for Technology Adoption, in: H.R. Shibly, A. Abdullah, M.W. Murad (Eds.), ERP Adoption in Organizations: The Factors in Technology Acceptance Among Employees, Springer International Publishing, Cham, 2022, pp. 103–148, https://doi.org/10.1007/978-3-031-11934-7_5.

[25] S. Pande, H.H.G. Savenije, A sociohydrological model for smallholder farmers in Maharashtra, India. Water Resour. Res. 52 (2016) 1923–1947.

[26] D. Djohan, J.J. Malard-Adam, S. Adla, S. Pande, Sociohydrological Model Structure Deficiency Assessment and Hybrid Model Selection, J. Hydrol. (2024). Under review.

[27] P. Steduto, T.C. Hsiao, D. Raes, E. Fereres, AquaCrop—The FAO Crop Model to Simulate Yield Response to Water: I. Concepts and Underlying Principles, Agron. J. 101 (2009) 426–437, https://doi.org/10.2134/agronj2008.0139s.

[28] D. Raes, P. Steduto, T.C. Hsiao, E. Fereres, AquaCrop—The FAO Crop Model to Simulate Yield Response to Water: II. Main Algorithms and Software Description, Agron. J. 101 (2009) 438–447, https://doi.org/10.2134/agronj2008.0140s.

[29] National Aeronautics and Space Administration, n.d. Prediction Of Worldwide Energy Resources (POWER) API. URL https://power.larc.nasa.gov/ (accessed 7.18.24).

[30] Agriculture Marketing, n.d. URL https://agmarknet.gov.in/default.aspx (accessed 7.18.24).

[31] Django REST framework, n.d. URL https://www.django-rest-framework.org/(accessed 7.18.24).

[32] S.I. Jiménez-Jiménez, W. Ojeda-Bustamante, M.A. Inzunza-Ibarra, M.de J. Marcial-Pablo, Analysis of the NASA-POWER system for estimating reference evapotranspiration in the Comarca Lagunera, Mexico. Ing. Agríc. Biosist. 13 (2021) 201–226, https://doi.org/10.5154/r.inagbi.2021.03.050.

[33] D. Raes, P. Steduto, T.C. Hsiao, E. Fereres, AquaCrop Version 6.0 –6.1, Reference Manual, Food and Agricultural Organization of the United Nations, Rome, Italy, 2018.

[34] G.H. Hargreaves, Z.A. Samani, Reference Crop Evapotranspiration from Temperature, Appl. Eng. Agric. 1 (1985) 96–99.

[35] S. Shahidian, R. Serralheiro, J. Serrano, J. Teixeira, N. Haie, F. Santos, Hargreaves and Other Reduced-Set Methods for Calculating Evapotranspiration, in: A. Irmak (Ed.), Evapotranspiration, IntechOpen, Rijeka, 2012, https://doi.org/10.5772/18059.

[36] Soil and Land Resources Assessment Division, Indian Space Research Organisation, 2016. National Information System for Climate and Environment Studies (NISCES).

[37] Bhattacharyya, T., Chandran, P., Ray, S.K., Mandal, C., Pal, D.K., Venugopalan, M. V., Durge, S.L., Srivastava, P., Dubey, P.N., Kamble, G.K., Sharma, R.P., Wani, S.P., Rego, T.J., Pathak, P., Ramesh, V., Manna, M.C., Sahrawat, K.L., 2008. Physical and Chemical Properties of Red and Black Soils of Selected Benchmark Spots for Carbon Sequestration Studies in Semi-Arid Tropics of India: Global Theme on Agroecosystems Report no. 35 [WWW Document]. URL https://oar.icrisat.org/2331/(accessed 7.17.24).

[38] J.G. Arnold, D.N. Moriasi, P.W. Gassman, K.C. Abbaspour, M.J. White, R. Srinivasan, C. Santhi, R.D. Harmel, A. Van Griensven, M.W. Van Liew, others, SWAT: Model use, calibration, and validation, Trans. ASABE 55 (2012) 1491–1508.

[39] C.K. Abbaspour, SWAT Calibrating and Uncertainty Programs. A User Manual, EAWAG, Zurich (Switzerland), 2015.

[40] A. van Griensven, T. Meixner, S. Grunwald, T. Bishop, M. Diluzio, R. Srinivasan, A global sensitivity analysis tool for the parameters of multi-variable catchment models, J. Hydrol. 324 (2006) 10–23.

[41] Q. Duan, S. Sorooshian, V. Gupta, Effective and efficient global optimization for conceptual rainfall-runoff models, Water Resour. Res. 28 (1992) 1015–1031, https://doi.org/10.1029/91WR02985.

[42] J.A. Nelder, R. Mead, A Simplex Method for Function Minimization, Comput. J. 7 (1965) 308–313, https://doi.org/10.1093/comjnl/7.4.308.

[43] J.H. Holland, Adaptation in Natural and Artificial Systems: Adaptation in Natural and Artificial Systems An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence, Bradford Books, MIT Press, Cambridge, MA, USA and London, UK, 1992.

[44] J.-H. Jeon, C.-G. Park, B.A. Engel, Comparison of Performance between Genetic Algorithm and SCE-UA for Calibration of SCS-CN Surface Runoff Simulation, Water 6 (2014) 3433–3456, https://doi.org/10.3390/w6113433.

[45] R.G. Allen, L.S. Pereira, D. Raes, M. Smith, Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56, Food and Agricultural Organization of the United Nations, Rome, 1998.
[46] E. Ekström, J. Halonen, Hydro-climatic Risk Assessment and Communication for Smallholder Farmers in Maharashtra (Bachelor Thesis), KTH School of Industrial Engineering and Management, Stockholm, 2021.
[47] J. Desloires, D. Ienco, A. Botrel, Out-of-year corn yield prediction at field-scale using Sentinel-2 satellite imagery and machine learning methods, Comput. Electron. Agric. 209 (2023) 107807, https://doi.org/10.1016/j.compag.2023.107807.
[48] A. Paliwal, Balwinder-Singh, S. Poonia, M. Jain, Using microsatellite data to estimate the persistence of field-level yield gaps and their drivers in smallholder systems, Sci. Rep. 13 (2023) 11170, https://doi.org/10.1038/s41598-023-37818-2.
[49] M. Jain, B. Singh, A.A.K. Srivastava, R.K. Malik, A.J. McDonald, D.B. Lobell, Using satellite data to identify the causes of and potential solutions for yield gaps in India's Wheat Belt, Environ. Res. Lett. 12 (2017) 094011, https://doi.org/10.1088/1748-9326/aa8228.
[50] M. Sussy, H. Ola, F.A.B. Maria, B.-O. Niklas, O. M Cecilia, O.-K. Willis, M. Håkan, G. Djurfeldt, Micro-Spatial Analysis of Maize Yield Gap Variability and Production Factors on Smallholder Farms, Agriculture 9 (2019) 219, https://doi.org/10.3390/agriculture9100219.
[51] A. Pelosi, H. Medina, P. Villani, G. D'Urso, G.B. Chirico, Probabilistic forecasting of reference evapotranspiration with a limited area ensemble prediction system, Agric. Water Manag. 178 (2016) 106–118, https://doi.org/10.1016/j.agwat.2016.09.015.