Agricultural Systems 187 (2021) 103014

Contents lists available at ScienceDirect

Agricultural Systems

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

Crop modelling in data-poor environments – A knowledge-informed probabilistic approach to appreciate risks and uncertainties in flood-based farming systems

Issoufou Liman Harou [a,b,*], Cory Whitney [c], James Kung'u [b], Eike Luedeling [c]

[a] World Agroforestry Centre (ICRAF), United Nations Avenue, Gigiri, P. O. Box 30677-00100, Nairobi, Kenya
[b] Department of Environmental Sciences, Kenyatta University, Kenya
[c] Department of Horticultural Sciences, University of Bonn, Auf dem Hügel 6, D-53121 Bonn, Germany

ARTICLE INFO

Keywords:
Complex systems
Crop model
Customized decision support tools
Limited information
Bayesian networks
Monte Carlo simulation

ABSTRACT

Crop models can support agricultural decisions, yet their reliability is necessarily limited when they do not sufficiently represent the complexity and specific circumstances of the target system. In some cases, models have such prohibitively high data requirements that they are only applicable with far-reaching and often questionable assumptions. In this paper, we demonstrate a customizable solution-oriented approach for crop modelling in situations where data and resources are limited. To address system complexity and produce a probabilistic crop model that does not depend on precise data, we used participatory analysis to describe system components using individual Bayesian networks that formalize expert knowledge into probabilistic causal relationships among important variables. We then used these Bayesian networks to generate inputs for a Monte Carlo model that illustrates the determinants of crop growth and simulates plausible ranges of expected grain and biomass yields at various stages of crop development. The resulting model accounts for all important variables and their interactions, as examined by local and foreign experts and described in relevant literature. We describe how to develop and customize such a model to specific situations based on case studies related to flood-based farming systems in Ethiopia and Kenya. The model assesses the performance of cropping systems and individual crops, and identifies factors of high importance for system outcomes. This approach to crop modelling paves the way for new opportunities to support agricultural decisions, since it does not require perfect information and can accommodate system complexity and uncertainty in data-poor environments.

1. Introduction

Various types of crop models are available with applications for both science and policy (Boote et al., 1996; Murthy, 2004; Uusitalo et al., 2015; Van Ittersum et al., 2013). While these models contribute to understanding the determinants of agricultural production in diverse contexts, their application in supporting decision-making is still confronted with many challenges. These challenges arise from conflicts between two distinct objectives of modelling: (1) summarize existing knowledge to formulate new research questions and (2) contribute actionable information to support real-life decisions aiming to produce desirable impacts (Boote et al., 1996; Luedeling and Shepherd, 2016; Uusitalo et al., 2015). Another major challenge for crop modelling is the need for frameworks that sufficiently address system complexity while allowing the use of imprecise or incomplete data inputs.

A crop model that seeks to generate knowledge does not necessarily have to fully capture the complexity of a system, but this is critically important for models that aim to provide realistic management advice (Boote et al., 1996; Uusitalo et al., 2015). For crop models that simplify the complexity of agricultural systems too much, application beyond the originally intended scope can easily result in unrealistic yield estimates, particularly when factors that are not represented in the model are important (Murthy, 2004; Uusitalo et al., 2015; Van Ittersum et al., 2013).

Whenever the purpose of a model is primarily to explore the dynamics of tightly circumscribed systems, extensive (and often costly)

* Corresponding author at: World Agroforestry Centre (ICRAF), United Nations Avenue, Gigiri, P. O. Box 30677-00100, Nairobi, Kenya.
E-mail addresses: issoufoul@gmail.com, limanissoufou@students.ku.ac.ke, L.issoufou@cgiar.org (I. Liman Harou), cory.whitney@uni-bonn.de (C. Whitney), kungu.james@ku.ac.ke (J. Kung'u), luedeling@uni-bonn.de (E. Luedeling).

https://doi.org/10.1016/j.agsy.2020.103014
Received 18 February 2020; Received in revised form 29 November 2020; Accepted 30 November 2020
Available online 21 December 2020
0308-521X/

data collection can address data limitations (Murthy, 2004; Refsgaard et al., 2007). For models that are meant to be applied beyond the scope of such heavily-sampled settings, however, data availability in the intended application areas should be considered when designing models. A mismatch between model needs and data supply can easily preclude the use of models in new settings or require the use of default values or other arbitrary choices for required model inputs for which no data are available. Crop models for supporting agricultural decisions must be developed with a view to the range of input data they depend on, since decisions being supported often require the inclusion of aspects on which little information is available (Luedeling and Shepherd, 2016). Consequently, such models would greatly benefit from being able to work with imprecise or incomplete datasets. They should also be able to account for the uncertainties that surround virtually all variables used in crop simulations. In order to specify precise numbers with reasonable confidence, any crop growth simulation would ideally be preceded by an extensive data gathering exercise, including crop growth experiments (Baroni and Tarantola, 2014). While hard data may be desirable and may often be more accurate than soft ones, time and resource constraints often preclude the collection of such precise datasets. In such situations, crop models that accommodate imperfect information would help to set realistic expectations about data inputs (Baroni and Tarantola, 2014; Luedeling et al., 2015).

The dual challenges of system complexity and data limitations highlight the need for new modelling approaches that convey holistic systems understanding and do not require perfect information. In this paper, we argue that crop models should consider data uncertainty and include diverse data types and sources. We propose a framework for meeting this challenge and demonstrate its usefulness by developing a flexible crop model for supporting the management of Flood-based Farming Systems.

Flood-based Farming Systems (FBFS) are rainfed agricultural systems found mostly in relatively low-lying areas that frequently experience flooding. While such floods are difficult to predict, they occur often enough – and are reliable enough – to allow the existence of farming systems that use this floodwater for irrigation, substantially extending the scope of crop production. Due to the critical importance of floodwater, which often constitutes a vital resource for substantial numbers of small farms, water acquisition and sharing among farmers are generally governed by complex socio-institutional arrangements at various scales (Haile, 2010; Van Steenbergen et al., 2010). In this context, floods are not considered natural hazards but simply manifestations of natural water fluctuation (e.g. periods of high rainfall or flood pulses in reservoirs), as described in the concept of flood pulse (Junk et al., 1989), or in the concept of Crue/Décrue (Harlan and Pasquereau, 1969). Water management in FBFS takes advantage of water surplus related to flood events, which is then stored to allow extension of the natural growing period.

Despite some scientific investigations, FBFS are still not well understood (Erkossa et al., 2014; FBLN, 2018; Harlan and Pasquereau, 1969; Liman Harou et al., 2020; Puertas et al., 2011; Van Steenbergen et al., 2011). Most FBFS studies have focused on specific issues such as moisture conservation, improvements to water diversion structures or social organisation (Embaye et al., 2012; Haile, 2010; Kilongosi et al., 2019). The lack of policy and systematic approaches have made evidence-based management of FBFS difficult (Haile, 2010; Puertas et al., 2011; Van Steenbergen et al., 2010, 2011). Many engineering headworks that worked well under conventional irrigation have failed when transferred to FBFS settings, as evidenced by multiple examples in the Oromia and Tigray regions of Ethiopia (Erkossa et al., 2014). While such failures are attributed to knowledge limitations regarding the timing, frequency and size of the flood for irrigation, the risks have not always been sufficiently appreciated due to a lack of adequate risk assessment procedures. Consequently, farmers, investors and donors cannot reliably assess the return on investment in FBFS. Given that FBFS have been shown to generate substantial benefits in many places compared to rainfed agriculture (e.g. Kobo, Ethiopia; Erkossa et al., 2014; Van Den Ham, 2008), this may have resulted in many lost opportunities. However, positive outcomes can only materialize when the additional water is well distributed to cover shortages in rainfall. Inadequate flood management can result in waterlogging of soils or submergence of crops and have negative impacts on crop production. Models aiming to make effective predictions for FBFS must consider various factors such as sediment management, hydraulic infrastructure and the social rules set by farmers (Liman Harou et al., 2020). Such factors are generally absent from current crop models. Consequently, there is, to our knowledge, no crop model that is suitable for FBFS. Therefore, the objective of this paper is to use the settings of FBFS to demonstrate the development of a crop model for complex agricultural systems with limited information and high uncertainty.

We applied methods from Decision Theory to simulate crop performance under FBFS with full consideration of production risks and inclusion of all available information and knowledge on important drivers of system behaviour. We used a mixed methods approach, incorporating both Bayesian Network (BN) and Monte Carlo (MC) models to develop a crop model that captures farmers' realities and accounts for qualitative as well as quantitative information. The BN is used to describe important qualitative processes, mostly related to farming constraints, such as the cropping systems and management options adopted by farmers. The MC model is used to describe the quantitative processes, such as biomass accumulation across crop development stages.

FBFS are used to showcase the methodology, yet the concept is customizable to any other complex system. It is particularly suitable for solution-oriented research. Two study areas in Ethiopia and Kenya were selected as reference locations to produce a generic model that could be applied to many situations and contexts with minimal modification. The two areas are assumed to cover much of the complexity (e.g. type of the biophysical system, agricultural management, social and institutional arrangements) found in FBFS. Data analysis was conducted using the R programming language (R Core Team, 2019) and software packages mentioned in the text are contributed packages of R.

2. Methods

2.1. Description of the study area and the sampling frame

The study areas, Kisumu County in Kenya and the Tigray region in Ethiopia, are located within relatively low-lying regions (Fig. 1C, D). These flood-prone areas differ substantially in terms of system hydrology, management of agronomic flooding and other agricultural practices (Fig. 1E, F; Table 1). While FBFS farmers in Kisumu mainly use permanent reservoirs to irrigate crops via inundation canals, farmers in Tigray mostly obtain water from ephemeral rivers, where they have to divert important amounts of spate flow within a relatively short period of time (Table 1). While the water sources and acquisition procedures vary among systems based on rice (*Oryza sativa)* in Kisumu (i.e. 'outgrowers', East and West Kano), the cultural practices are similar. In these areas, farmers sow rice in monoculture during floods, followed by various flood recession crops that are intercropped depending on farmers' preferences and water availability (Fig. 1E). In Tigray, in contrast, farmers mainly sow maize (*Zea mays*), sorghum (*Sorghum bicolor*) and tef (*Eragrostis tef*) based on rainfall to later irrigate them using flood water (Fig. 1F). The sampling frame of the study considered 8 areas across Kisumu and Tigray to capture such important differences (Table 1):

- East and West Kano (Kisumu): Two conventional irrigation schemes where the Kenyan national irrigation board uses large pumps to provide farmers with water from River Nyando and Lake Victoria. We considered these locations because they share many properties with FBFS while representing a special case of water acquisition.

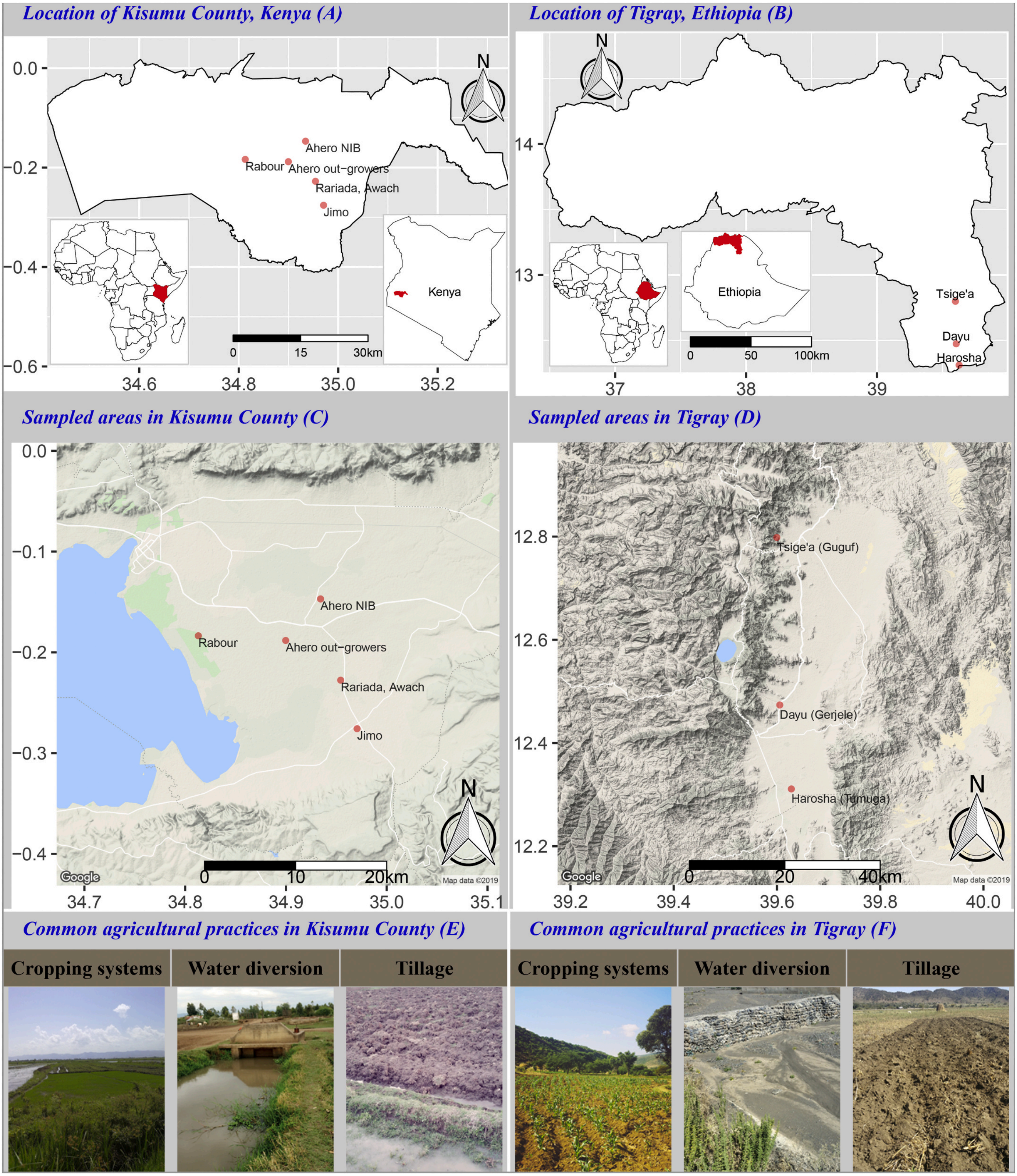

Fig. 1. Description of the sampling frame used to develop a mixed model of various flood-based farming practices in the Tigray region of Ethiopia and Kisumu County in Kenya. The selection of the flood-based farming system schemes was based on field observations and recommendations from key informants. A) and B) indicate the locations of Kenya and Ethiopia in Africa, the location of Kisumu County in Kenya and the location of Tigray in Ethiopia, and the sampling frame in both regions. For both areas, C) and D) show the prevailing topography around the sampling frame, and E) and F) provide a pictorial view of the local flood-based agriculture practices.

Table 1
Characteristics of the flood-based farming systems sampled to develop a crop model in the Tigray region of Ethiopia and Kisumu County in Kenya.

Study region	Sampled areas	Original design idea	Diversion type	Water Source	Water Source hydrology	Water acquisition
Kisumu	East Kano	Engineers	Modern	Nyando River	Permanent	Pump
	West Kano	Engineers	Modern	Lake Victoria	Permanent	Pump
	Ahero out-growers	Farmers	Traditional	Nyando River, East Kano	Permanent	Gravity
	Awach out-growers	Farmers	Traditional	Awach river	Permanent	Gravity
	East Nyankach	Farmers	Modern and Traditional	Runoff harvesting	Ephemeral	Household pond/roof
Tigray	Tsige'a (Guguf)	Farmers and engineers	Improved	Dry wadis	Ephemeral	Gravity
	Dayu (Gerjele)	Engineers	Modern	Dry wadis	Ephemeral	Gravity
	Harosha (Tumuga)	Farmers	Traditional	Dry wadis	Ephemeral	Graviy

Agricultural practices are mainly homogeneous within study regions. The term 'out-growers' refers to farmers outside the scope of the Kenyan national irrigation board. Traditional flood water diversions are physical infrastructure, such as deflecting spurs or soil bunds that are constructed by farmers across flood channels using locally available materials. Modern diversion structures, such as diversion weirs, are usually designed by engineers and made of concrete. The improved diversion type constitutes an integration of farmers' and engineers' knowledge.

- Ahero and Awach out-growers (Kisumu): Two FBFS schemes where farmers use simple gravity flow to acquire water from the Nyando and Awach Rivers. The main difference between these locations is that the Ahero out-growers scheme serves as safe disposal for excess water in East Kano.
- East Nyankach (Kisumu): Scheme where farmers collect water via runoff and roof water harvesting. Typically, they store the collected water in various household ponds and water tanks to be used later for irrigation.
- Tsige'a, Harosha, and Dayu (Tigray): Schemes where farmers use improved, traditional, and modern floodwater diversions, respectively, to harness spate flow from various dry wadis.

The differences captured in the 8 areas considered in the sampling frame (Fig. 1E, F; Table 1) have implications for water supply, social organisation, agricultural management and crop development. These aspects translate into different farmer concerns in the two study regions. The most important concerns, which vary across the study regions, are related to water supply and agricultural management. While floodwater uncertainty seems to be more important for farmers in Tigray than for those in Kisumu County, the challenges related to agricultural management appear more pressing in Kisumu County than in Tigray. In both areas, weeds are important constraints and weed removal is required at every stage of crop development.

2.2. Participatory model development

To describe crop performance within FBFS settings, we developed a mixed Bayesian Network/Monte Carlo model describing causal relationships deemed important for crop production in FBFS. We parameterized the model using various sources of information, including literature review, online databases, expert knowledge elicitation, farmer and expert interviews and remote sensing (Krueger et al., 2012; Liman Harou et al., 2020; Refsgaard et al., 2007). The goal was to develop a detailed description of the system by breaking down the system complexity into all important processes (see Section 2.5) while highlighting risk factors that influence crop production and accounting for data uncertainty (Refsgaard et al., 2007). To ensure the quality of the estimates, we introduced the experts to the principles of decision analysis and provided them with a calibration training (Hubbard, 2014a; Luedeling et al., 2015; Whitney et al., 2018a, 2018b). Expert calibration is the process of training experts to provide robust estimates and to assess their own confidence in these estimates. The procedure consists in training the experts in estimation, so that they neither understate nor overstate their knowledge of the issue under consideration. We conducted the calibration training based on a set of predefined questions guiding the experts as recommended by Hubbard (2014a, 2014b). Typically, the procedure consists in asking the experts to estimate the values of some variables based on a set of trivia questions and assessing their skill while providing tools for making better estimates. We do several rounds of this until the experts' skills have increased to an acceptable level (i.e. they are correct close to 90% of the time when they say they have 90% confidence in their estimate).

The model development process, therefore, consisted of five sequential steps with specific milestones (Fig. 2) to develop theories that match farmers' realities:

- A literature review to understand the issue and design broad leading questions to primary experts (experts from various academic and research institutions working with FBFS and related fields).
- A high-level discussion with 11 primary experts to understand important FBFS concepts. At this stage of the process, the main objective was to obtain a high-level understanding of the concept of FBFS from different perspectives. We selected the experts based on their relevant and extensive experience regarding the practice of FBFS (Krueger et al., 2012) and asked them to define FBFS, thereby identifying important concepts to include in the model.
- Discussions with 20 focus groups in Ahero and Kisumu towns (Kenya) in December 2016 and June 2017 and in Mekelle and Alamata towns (Ethiopia) in December 2016 and January 2017. At this stage of model development, the objective was to relate the high-level concepts to the local contexts. This allowed us to develop specific understanding of the FBFS practices in both Kisumu County and Tigray.
- Consultation with local experts to formalize the model. We extended the pool of local experts, initially represented by local farmers and extension civil servants, to include the participants of a leadership course in flood-based farming and water harvesting in Kenya and the participants of the International Training Course on Integrated Watershed Management and FBFS in Ethiopia (FBLN, 2018). The main objective of this expert consultation was to define the causal relationships between variables along with the variable estimates. To achieve this objective, we grouped the farming constraints into variables related to soil water (e.g. upstream abstraction, amount of floodwater reaching the farming plot, available soil water), cropping systems (e.g. crop type, previous crop, effectiveness of cropping systems) and farmer management (e.g. access to inputs, skill of the farmer, pest and disease impacts). At this stage, we shared the information acquired in the previous steps with the local experts for their feedback. They helped us to formalize the model and link the qualitative variables using a BN, and the quantitative ones using an MC model (Figs. 2 and 3). We later connected the BN and the MC models to form a generic model that captures the variability of FBFS schemes in the study regions.
- Interviews with 159 farmers to ensure that the model had captured farmers' realities. We asked the farmers to estimate their historical yields at 90% confidence intervals and used these estimates along with the experts' estimates (e.g. yield potential) to model important yield metrics (e.g. yield gap).

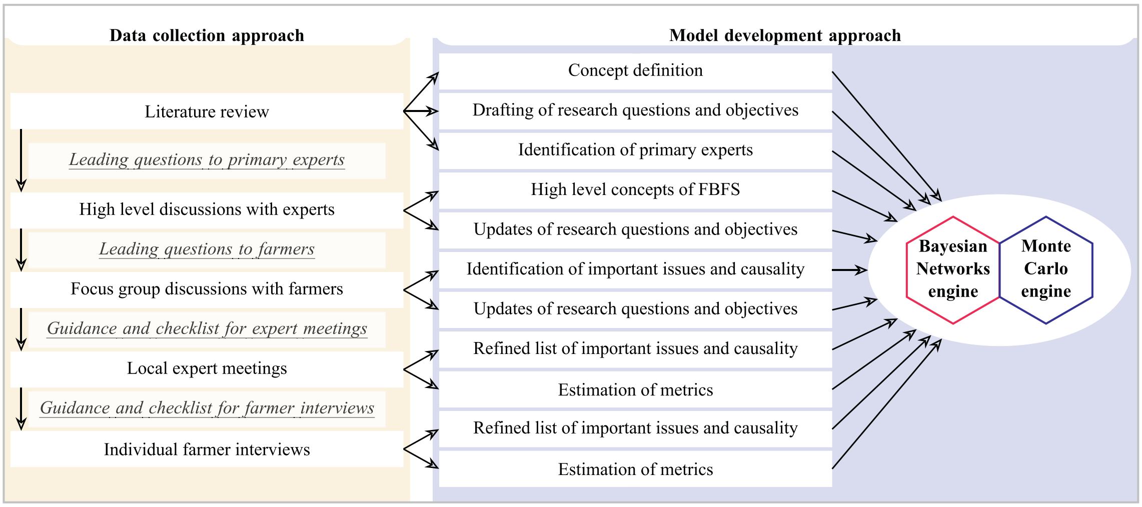

Fig. 2. Overview of the approach used to develop a crop model for flood-based farming systems in Tigray, Ethiopia, and Kisumu County, Kenya.

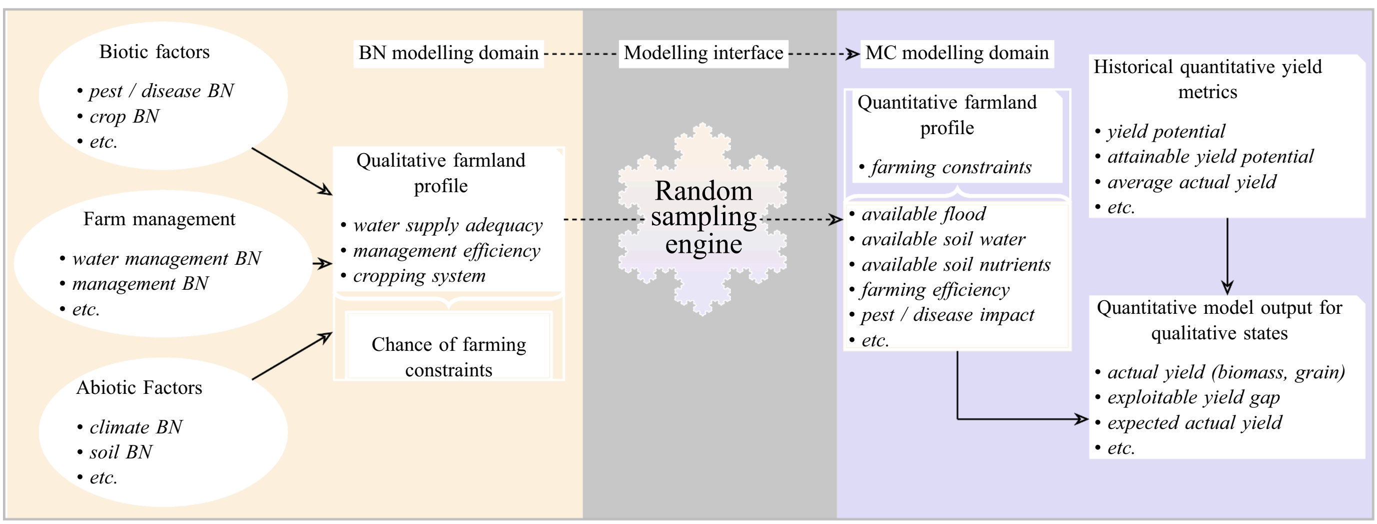

Fig. 3. Conceptual framework of important processes considered in the development of a crop model for flood-based farming systems in the Tigray region of Ethiopia and Kisumu County in Kenya.

2.3. Models for quantitative variables

A Monte Carlo (MC) simulation is a quantitative procedure that uses a deterministic mathematical equation to estimate the values of continuous variables based on the probability distributions of one or more input variables. The population of outputs consists of drawing many sets of random values from these distributions to parameterize the central equation, hence expressing a reasonable estimate of uncertainty around the value of the target variable (Luedeling et al., 2015; Refsgaard et al., 2007; Rosenstock et al., 2014).

We used an MC model with 21 variables determining the performance of a given farmland (e.g. yield potential, exploitable yield potential, exploitable yield gap, exploitable yield gap loss due to constraints, actually exploited yield gap, expected actual yield, farming constraints and crop growth estimation considered at different stages of crop development) to quantitatively describe biomass accumulation during crop development and the conversion of this biomass into grain. We estimated the yield potential based on the global yield gap database (Van Ittersum et al., 2013), the attainable yield potential based on the literature (Van Ittersum et al., 2013), the actual yield based on data from farmer interviews, and the remaining variables based on expert elicitation. For each variable, the input data are the 5% and 95% quantile values, along with the shape of the probability distribution. We considered the yield potential to be the theoretical yield limit of a variety grown under optimal conditions (Van Ittersum et al., 2013). Since achieving the full yield potential is unrealistic, we considered the exploitable yield potential as the fraction of yield potential (70–80%) that farmers can potentially achieve under ideal conditions. Since such ideal conditions are not encountered in most places, we considered the exploitable yield gap (difference between the exploitable yield potential and the average farmer yield) to account for the systemic constraints encountered on farms. Internally, we imposed restrictions on some of the correlated variables to avoid simulating inconsistent values. For example, we restricted the exploitable yield gap to non-negative values that never exceed the attainable yield values in the Monte Carlo simulation. Finally, the expected actual yield expressed the yield expectation given a farmer's management conditions (Van Ittersum et al., 2013).

2.4. Models for qualitative variables

A Bayesian network (BN) is a multivariate cause-effect model suitable for conditional reasoning across variables with multiple qualitative states. Typically, a BN defines the probabilities of the states of unknown discrete variables, conditional on the probabilities of the states of (an) other known variable(s). These conditional probabilities express causal relationships between the discrete variables (Jensen, 1996; Pearl, 1988). A BN is composed of logical connections (directed edges) between variables (nodes), which are arranged in probabilistic graphical models. Dependencies are represented by arrows (arcs) encoding the direction and nature (i.e. direct or indirect) of the causal relationships (Pearl, 1988). A node describes the states of a variable (node states), with a variable at the arrowhead (child node) being influenced by the variable at the arrow tail (parent node). The strength of the influence is portrayed by the conditional probabilities typically specified as stochastic matrices known as conditional probability tables (CPTs), which are the central elements of BNs (Fenton and Neil, 2013; Jensen, 1996; Pearl, 1988). They portray the strength of the association between all states of a child node and the various states of its parents (Hansson and Sjökvist, 2003; Scutari and Denis, 2015).

We assumed that management practices influence various biotic and abiotic factors which, in turn, drive system functions (Jax, 2005). These profile a given farmland in terms of farming constraints and performance. The variability among system processes across farmlands is considered to imply differences in the states of variables modulating them. In developing the BN, the local experts used 121 variables to specify complex causalities defining one important variable: the 'farming constraints' factor, which accounts for the combined effects of all limiting factors. The 121 variables describe the adequacy of water supply, cropping systems and agricultural management (see supplementary code for more details on these variables). The experts disaggregated these qualitative variables into their respective states (e.g. it rains or it does not, manure is either applied or it is not) and estimated several parameters describing the probabilities of these states as CPTs (see Section 2.6).

2.5. Modularity and system complexity

The generic model encapsulates the complexity of the farming system at 4 levels of abstraction embedded in the BN and MC models:

- Recognition of individual processes defining system performance: This was done in plenary sessions during workshops, where the experts identified the important processes and grouped them into separate modules based on the assumption that farming constraints affect crop development. To facilitate the description of complex processes, we described some of these modules using submodules. The experts described these processes through causal chains describing resource allocation (e.g. available soil water or available soil nutrients) in three modules (the soil water, the cropping system and the management modules) to estimate the farming constraints variable via the BN. Then, the experts assessed crop growth under these constraints at various stages of crop development, which we used to inform the MC model.
- Disaggregation of the processes into individual variables (e.g. rainfall occurrence or manure application): This was done in working groups, with the aim of identifying important variables and interfaces for shared connections between the processes.
- Description of the variables: This was also done in working groups, with the aim of defining the relationships between variables and estimating their values. As mentioned earlier, the experts estimated the quantitative variables as continuous probability distributions and the qualitative ones as discrete probability distributions. When programming the model we used the *mcSimulation* function from the *decisionSupport* package (Luedeling et al., 2019) to conduct the MC simulation and the *cptable* function from the *gRain* package (Højsgaard, 2012) to generate the CPTs from the parameters estimated by the experts. For quantitative variables, we used the 5% and 95% quantile values along with the type of the distribution to construct estimates using the *estimate* function from the *decisionSupport* package, as required by the *mcSimulation* function. For qualitative variables, we used the *make_CPT* function from the *decisionSupport* package to formulate CPTs as required by the *cptable* function from the *gRain* package. The full BN and MC models describing these are provided in the technical and supplementary materials.
- Separation of crop development into critical stages: we defined the crop development stages from the onset of the cropping season to 10% ground cover as initial stage, from 10% ground cover to effective full cover as the development stage, from effective full cover to the start of maturity as the mid stage, and from the start of maturity to harvest or full senescence as the late stage (Allen et al., 1998). As mentioned earlier, we assumed that farming constraints can have effects on crops during each of the 4 development stages, and these effects depend on the state of farming constraints at the previous stage. For example, the effect of weeds due to inadequate weed removal during the initial stage can influence crop growth during the development stage. At each of these stages, we introduced a variable to measure crop growth and prevent crops from arbitrarily growing beyond the actual yield expectations. We assumed that crop yield follows a negatively skewed gamma distribution (Gallagher, 1987; Ramirez et al., 2003), meaning that there is a high chance of low yield and relatively low chance of high yield. In developing the model, we focused on 4 crops – rice, maize, sorghum and tef – based on their importance in the study areas.

2.6. Modelling interface

As mentioned earlier, the experts estimated several parameters describing joint CPTs as inputs for the *make_CPT* function. To create the full CPT, the function requires the prior probability distribution of the child node, the child node's sensitivity relative to the parents, the parents' effects, and the weight of influence of each parent (Hansson and Sjökvist, 2003; Luedeling et al., 2019; Whitney et al., 2018b). We automated the *make_CPT* routines to generate inputs for the *cptable* function from the *gRain* package (Højsgaard, 2012) and formalised the BN as a computer-readable graphical model following the experts' causal reasoning (Liman Harou et al., 2020).

While BNs are ideal for specifying complex levels of hierarchy and categorization as cause-effect relationships, they also provide a convenient interface for studying simple aspects based on the states of different variables and their causal linkages. Based on the different combinations of the states of its parents, we generated the probabilities of the 'farming constraints' node by sampling from the posterior distributions of the BN (Liman Harou et al., 2020). This is equivalent to generating all possible farming constraint scenarios based on the BN model. Technically, we used MC particle filters (Kitagawa, 1996; Koller and Friedman, 2009; Scutari, 2010) to generate the probabilities of farming constraints conditional on different combinations of the states of variables involved in the BN. We assumed that the negative effects of the different states of the variable on crop development range between 0 and 1, where 0 denotes ideal conditions (exploitable yield potential) and 1 the worst-case scenario (total loss). Therefore, we used the variable as expected loss ratio, accounting for both the effects of the variable states and their probability of occurrence in the MC model (Fig. 3).

A combination of MC and BN models is ideal for specifying complex deterministic and cause-effect relationships while enabling the use of data and information that are not necessarily precise. These modelling approaches can produce reliable advice that makes full use of the current state of information and adequately expresses all uncertainties in such a way that outcomes and risks of a given decision can be appreciated. We set the MC model to simulate the various yield metrics (Fig. 3) ranging from the maximum possible yield to what farmers actually achieve (Van Ittersum et al., 2013; Liman Harou et al., 2020), and the BN to provide the shape of the probability distributions of the variable 'farming constraints' (Fig. 3), as well as its 90% confidence interval and median values. This was done at each of the 4 crop development stages to account for the variability of farming constraints over the growing season and the biomass accumulation over time.

Note that we evaluated the probability distributions of the node 'farming constraints' based on reasonable bounds for skewness and kurtosis via visual observations supported by bootstrapping. We fitted several candidate distributions using the *fitdist* function from the *fitdistrplus* package (Delignette-Muller and Dutang, 2015; Liman Harou et al., 2020), from which we chose the best-fitting distribution. In the supplementary materials, we show examples of Cullen and Frey graphs, and quantile and probability plots, along with the theoretical and empirical density and cumulative distribution functions used to facilitate the choice of the distributions. Note also that we computed the remaining parameters required by the *mcSimulation* function from the fitted distributions using the *fitdist* function and other functions from the *rriskDistributions* package (Belgorodski et al., 2017).

2.7. Inputs, outputs and special models

To demonstrate the flexibility of the approach and the importance of modularity in customizing model behaviour for specific situations, we derived specific models from the generic model to present the different modules and their throughputs in three case studies of varying complexity:

- The assessment of soil water using the soil water module looks at the challenge of maximizing water storage in soils, while limiting waterlogging and restricting water losses through evaporation, runoff or percolation in FBFS. We show how the model can be customized to prescribe optimum pre-season cultural practices for improved soil water on specific soil types. We formulated 18 different queries to assess the probability of soil water states (i.e. Drought risk, Normal, Waterlogging risk) conditional on soil types (i.e. Sandy, Loamy, Clayey), the amount of floodwater reaching the plot (Too little, Desired, Too much) and the states of manure application (True, False). We generated 10 variants of each of these 18 queries and ran them 1000 times using a customized version of the *cpdist* function from the *bnlearn* package (Scutari, 2010; Liman Harou et al., 2020). This corresponds to 180,000 possibilities capturing the variability of farmlands in space and time. It is important to note that by replicating each query 10 times we account for the variability of the same farmland. This is the variability over the same farming plot between different years, which technically corresponds to sampling from the posterior distribution of the BN. By considering each replication 1000 times, we account for the variability between different farmlands.
- The probabilistic assessment of crop biomass uses the crop growth module to show the model's suitability as monitoring tool for biomass accumulation (i.e. the amount of above-ground biomass produced by crops on a given farm) over the growing season. We considered all possible scenarios defining the likelihood of a given state of the variable 'farming constraints' to generate 81 different queries. These correspond to all possible combinations of the states of the variable 'farming constraints' taken at each of the 4 stages of crop development. We performed the queries using the *cpdist* function and used the outcomes in a Monte Carlo simulation to generate 10,000 model runs simulating the exploitable yield gap, the exploitable yield gap loss due to constraints, the actual exploited yield gap, and the expected actual yield. While we provided all the 324 simulated results in the technical material, we discuss only a few of these here to showcase the model.
- The study of the impact of soil water and biotic stresses (i.e. pests, diseases and weeds) shows the effects of these on the grain yield of rice and sorghum, integrating all 4 model modules. We formulated all possible combinations of the extreme states of the nodes 'Soil water', 'Weed impacts' and 'Pest and disease impacts' based on which we queried the posterior distribution of the 'farming constraints' node. It is worth mentioning that we considered these scenarios only at the mid stage of crop development and assumed that the farming constraints define the actual yields, which the farmers estimated as value ranges. We conducted the MC simulation based on these considerations using the *mcSimulation* function.

It is important to note that inputs for the three case studies consisted of scenarios that were formulated based on child-parent state relationships. Outputs are the probability of adequate soil water, biomass and grain yield, respectively. The rationales of each of these case studies are provided in the supplementary materials.

3. Results

3.1. Farming constraints for FBFS in Kisumu County

Uncertainty in water supply in the sampled FBFS of Kisumu County depends mainly on the source of water and the water extraction method. Schemes relying on runoff harvesting are more exposed to water shortages than those relying on water bodies. For the latter type, seasonal fluctuations of water levels in reservoirs caused by the local hydrology can strongly affect water supply to FBFS schemes. Schemes relying on gravity are generally more exposed to water shortages than schemes equipped with pumps, particularly in periods of low water levels in the reservoirs. Risks related to water supply may be more prominent in the schemes of 'out-growers', which are exposed to water excess and shortages, due to the relatively low level of control on the amount of water reaching the plots, compared to the schemes managed by the Kenyan national irrigation board, where the water supply is more closely regulated.

In the schemes of 'out-growers', cultivation of a recession crop, which is almost entirely reliant on residual moisture, is not always possible due to lack of water. The risk of water excess is more of a concern for farmers in the Ahero out-growers scheme than for out-growers in Awach, where the risk of water shortages is more prominent. Despite being equipped with pumping machinery, the East Kano scheme is more exposed to water shortage than the West Kano scheme due to water level fluctuations in the Nyando River. The most important constraints related to agricultural management are represented by a range of pests, among which birds and other wild animals cause the greatest damage. Their impacts on crops are inversely correlated with water supply, meaning that the schemes with higher water endowments are more strongly exposed, because these animals often live close to water bodies. Relatively minor issues mentioned by farmers are the failure of the pumping machines in the East and West Kano schemes, the relatively limited labour force and access to inputs in the schemes of the 'out-growers'.

3.2. Farming constraints for FBFS in Tigray

Flood water uncertainty, which is perhaps the most important concern for farmers in Tigray, includes constraints related to the availability and management of floodwater. The performance of spate irrigation systems is generally assessed in terms of sediment budget, which strongly influences water delivery. The main challenge in these spate irrigation schemes is the sediment load and the labour-intensive work of sediment removal, which must be done regularly to maintain the flooding structures, particularly in traditional and modern floodwater diversion schemes. While fine sediments improve soil fertility, coarse sediments can raise wadi beds to levels that prevent water from reaching agricultural fields. In this regard, improved diversion structures that manage to filter out coarse sediments while allowing fine sediments to reach the farming plots, convey clear benefits to farmers (Embaye et al., 2012). In Tigray, floods often come from elsewhere and flow for a relatively short period. Farmers can easily miss such flood events, and large floods can cause damage to both infrastructure and crops.

Relatively minor issues mentioned by the farmers are limited access to inputs, particularly fertilizers and chemicals for controlling pests and diseases.

3.3. Overview of the conceptual model

The generic model (overview in Fig. 4) includes several socio-economic, biophysical, and agronomic aspects to describe crop development under various types of farming constraints. Socio-economic aspects in the model describe negotiations at various decisional scales (e.g. arrangements between upstream and downstream farmers), the socio-economic context (e.g. access to inputs, mutual assistance, available labour force, economic situation of the farmer). Agronomic aspects in the model describe farmers' choices regarding cropping systems and agricultural management. These include aspects such as crop type, crop variety, and the planting method along with other farm management practices such as the planting date or crop rotation. Aspects of the model related to farmer management inform soil fertility, crop protection and ultimately the efficiency of agricultural management. Biophysical aspects in the model describe the climate, the soil and the type of flooding system, which strongly influences floodwater delivery.

3.4. Soil water module

The soil water module (Fig. 4a) estimates the adequacy of water supply to the crop as a convenient determinant of crop performance (in the sense of water-limited yield). Adequacy is related to available soil water, which depends on other factors such as rainfall, soil water holding capacity, the amount of floodwater reaching the plot, and evaporation (Fig. 5). Two sub-modules describe this module: the amount of floodwater reaching the plot, and the available soil water (see supplementary material). The former describes external influences, infrastructure, and social arrangements determining flood water acquisition and sharing in FBFS. The latter describes the factor 'soil water', as it would be described under rainfed conditions (Fig. 5).

Various social arrangements, such as agreements for water sharing and system maintenance between stakeholders, determine the sharing of floodwater in a given scheme. These can influence sediment loads (depending on the type of water diversion), hence affecting the adequacy of water supply to crops through floodwater delivery. External effects such as upstream abstraction influence the amount of floodwater reaching the plot directly or indirectly (via the amount of shared flood water). Off-site rainfall occurrence can have a positive effect by increasing water supply. These, however, are not the only factors affecting soil water availability, which also depends on intrinsic soil characteristics, farm management, and other climatic factors (Fig. 5).

The assessment of soil water using the soil water module (Fig. 6) shows that on clayey soils without manure application, drought impacts are unlikely, while waterlogging is very likely when at least the desired amount of floodwater is obtained (Fig. 6). Uncertainty is greater when little floodwater is available. Supplementary manure application slightly increases the chance of waterlogging, which also increases when too much floodwater is obtained.

Loamy soils generally behave similarly to clayey ones, with a slightly

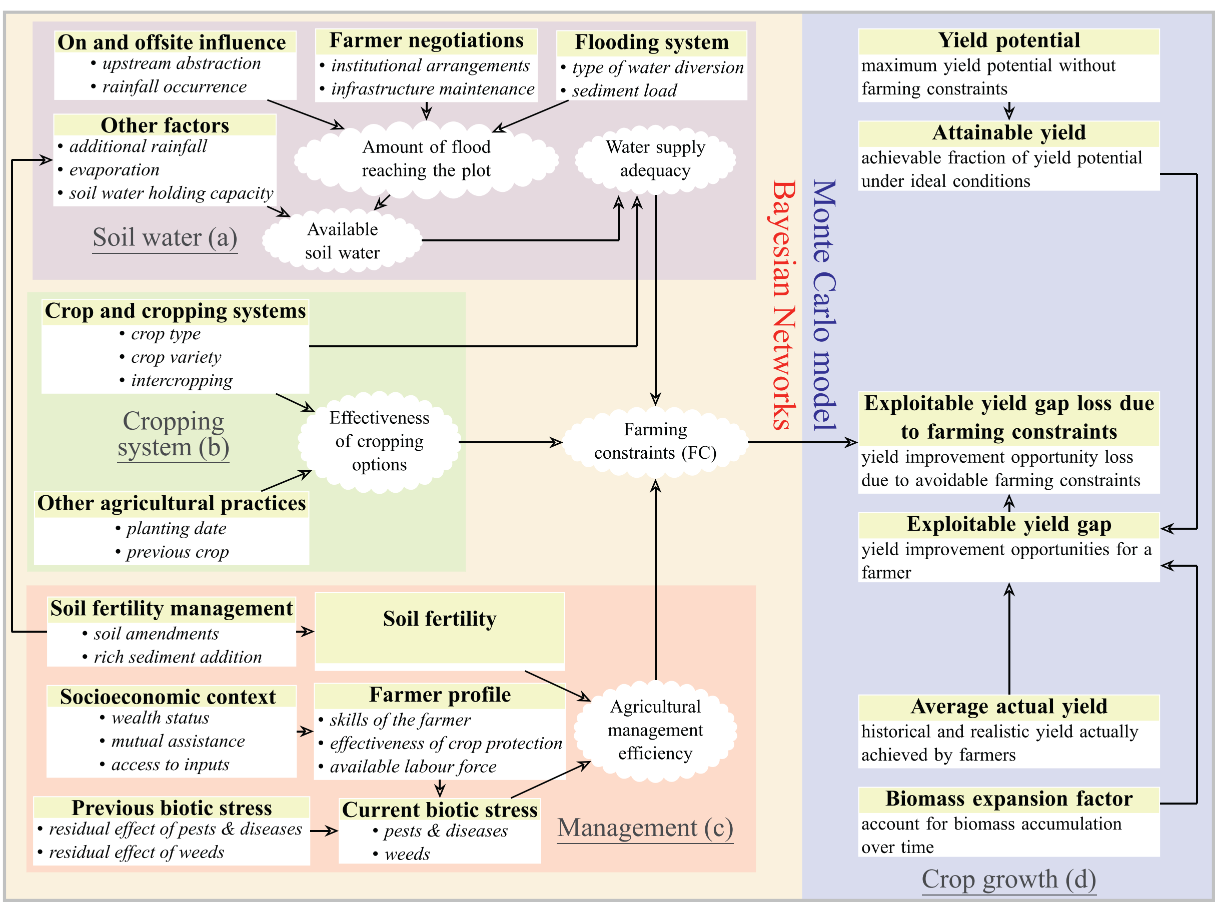

Fig. 4. Overview of the mixed Bayesian Network/Monte Carlo crop model for flood-based farming systems in the Tigray region of Ethiopia and Kisumu County in Kenya. a) factors determining the available soil water, b) cropping system characteristics, c) management practices adopted by the farmer, and d) crop development over time. This graph provides a simplified overview of the generic model. Nodes represent groupings of similar aspects. Bullet lists provide examples of variables that can be expected in the model described in fine detail in the technical material. Bubble shapes are Bayesian networks formed by variables (rectangular shapes) pointing to them.

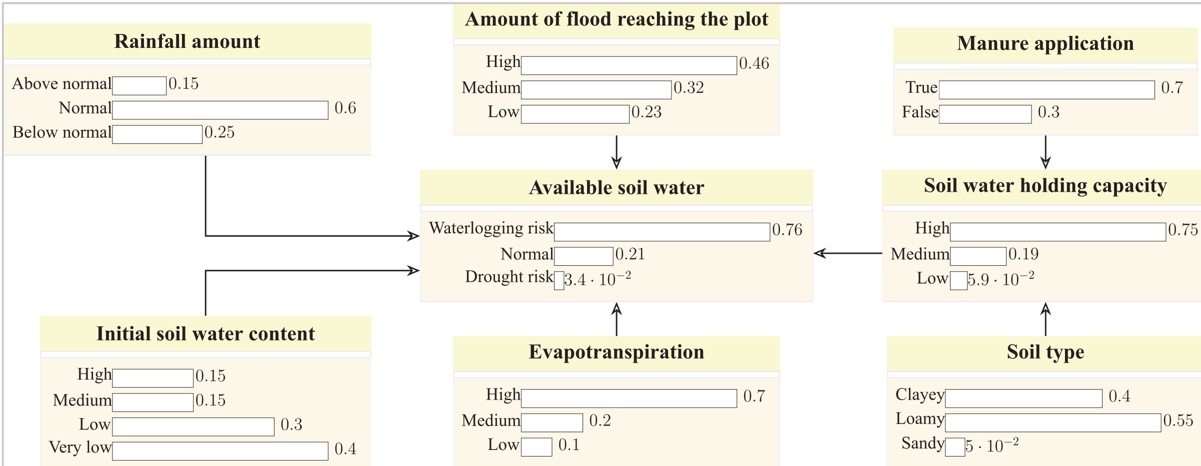

Fig. 5. Causal linkages defining the factor 'available soil water content' in flood-based farming systems in the Tigray region of Ethiopia and Kisumu County in Kenya (simplified illustration).

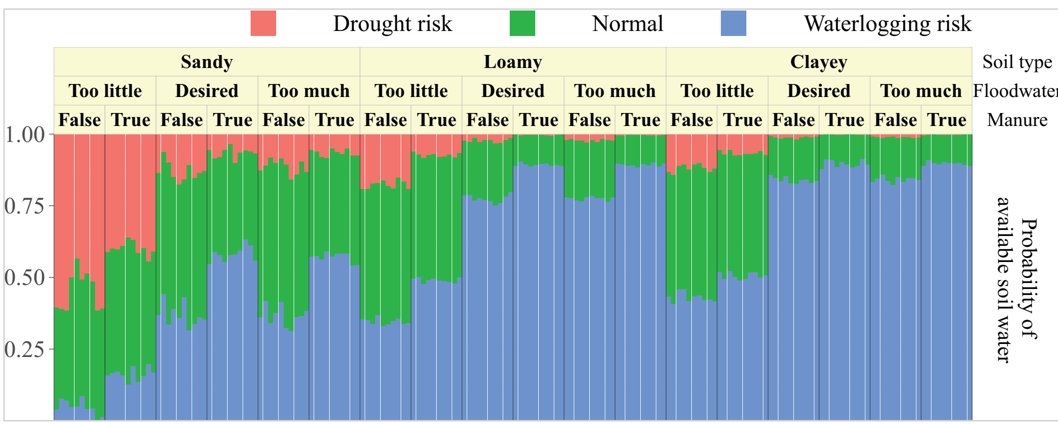

Fig. 6. Effect of soil type, manure and flood water on available soil water in flood-based farming systems in the Tigray region of Ethiopia and Kisumu County in Kenya. Each block of bars represents 10 possible outcomes of the corresponding query represented by the combination of the states of the 3 variables. Each of these 18 queries was run 1000 times resulting in a total of 180,000 model runs.

lower chance of waterlogging and greater probability of normal soil water conditions. Sandy soils appear to be the riskiest and most uncertain soil type for soil water. In all three soil types, drought appears to be more likely in years with little floodwater and no manure application than in other scenarios.

3.5. Cropping system module

The cropping system module (Fig. 4b) assesses the effectiveness of the cropping options, evaluating the performance of the cropping system adopted by the farmer instead of the cropping system itself. Crop types were ranked in decreasing order of yield (rice, tef, maize and sorghum) with improved crop varieties assumed to perform better than local ones. However, the effectiveness of the cropping system depends on an array of agricultural practices, such as the choice of planting date, the crops previously grown on the same land, and the presence and density of intercropping. The most effective cropping is generally achieved when an improved variety is planted early on land that had a different crop in the previous season. This can depend on the crop type. For example, rice and tef are quite tolerant of repeated cultivation and monoculture, whereas maize and sorghum perform better in crop rotations and intercropping.

3.6. Crop growth module

The crop growth module (Figs. 4d and 7) monitors biomass accumulation over time to estimate yield metrics (e.g. biomass yield gap, exploited biomass yield) at different stages of crop development. Fig. 7 summarises the MC model, including the relationships between the quantitative variables and the mathematical calculations involved at the initial and development stages. Crops are initially regarded in terms of their boundary conditions represented by their yield potential, which was differentiated into attainable yield potential, yield gaps, and actual yield via the actual farm conditions represented by farming constraints (see Section 2.3; Fig. 7). We selected the initial and the late stages of crop development along with the worst, medium, and best-case scenarios to show the biomass accumulation over time, and to illustrate the effect of varying farming constraints on crop development (Fig. 8). In this lumped assessment, we focused on the amount of biomass with no attempt to distinguish between the 4 crops considered.

The probabilistic simulation of crop biomass shows that several risk factors, some of which are specific to FBFS, affect the actual biomass yield, as well as the other yield metrics (Fig. 8). Farms generally have a high probability of relatively low yield, which varies substantially over time and in response to differing levels of farming constraints.

Based on computed 90% confidence intervals, the exploitable yield gap, the actually exploited yield gap, and the expected yield can differ

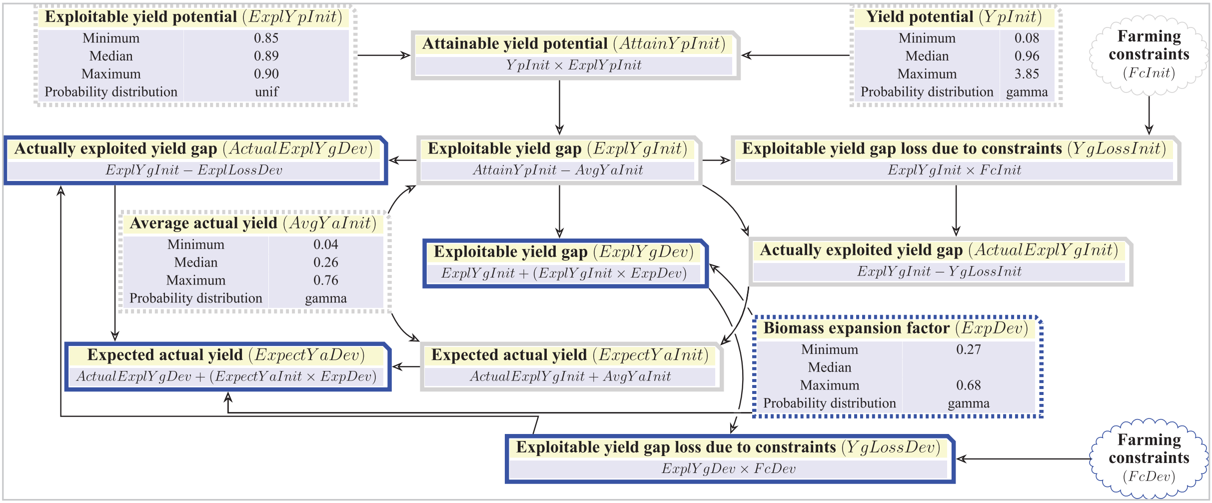

Fig. 7. Overview of a Monte Carlo model assessing biomass accumulation during crop development in flood-based farming systems in the Tigray region of Ethiopia and Kisumu County in Kenya. Gray-bordered nodes are taken at initial stage. Blue-bordered nodes are taken at development stage. Dashed and plain lines nodes, respectively, are elicited and computed. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

by a factor of two between situations with high and low farming constraints. The exploitable yield gap and loss due to farming constraints are substantial, indicating great potential for yield improvement. Exploitable yield loss is positively correlated with farming constraints, with a high chance of zero loss under minimal farming constraints. Simulation results show right-skewed gamma distributions (Fig. 8) indicating that crop yields in the study area are characterized by a relatively high frequency of low values. This highlights important uncertainty regarding yield expectations in the study areas. In general, chances of grain yield above 5 Mg ha^{-1} are negligible.

3.7. Management module

The management module (Fig. 4c) points out the agricultural management efficiency via 4 important sub-modules describing the general household situation, soil nutrients, pests and diseases, and weeds (Figs. 4c and 9). Pest and disease attacks can weaken the crop in subsequent development stages, aggravating other problems and ultimately lowering the agricultural management efficiency, particularly when the farmer lacks advanced skills in crop protection. The skill of the farmer, which depends on the social and economic context (e.g. opportunity, willingness and ability to purchase inputs), defines the farm management efficiency regarding water and crop protection methods (Fig. 4c).

The impacts of soil water and biotic stresses on grain yield suggest that crops subjected to strong farming constraints during early growth are likely to experience slow growth over the entire season. This cumulative effect of farming constraints seems to cause important biomass losses, which translate into low yields. Grain yield is very low when the late-stage constraints are high, regardless of the status of pests and diseases or water at the previous stages. Under such conditions, grain yield in both rice and sorghum rarely exceeds 1 Mg ha^{-1}. Grain yield expectations are generally higher and more sensitive to varying levels of farming constraints in sorghum compared to rice. The greater sensitivity of sorghum can be seen by comparing the two crops following increasing levels of farming constraints (from bottom to top in Figs. 10 and 11). Based on the confidence interval, grain yield of sorghum surpasses that of rice by about 1 Mg ha^{-1} when the farming constraints are low at the initial and development stages. This difference is on the order of 0.5 Mg ha^{-1} when the farming constraints are high at the initial and development stages. These findings highlight the importance of the early and late stage regarding grain yield in FBFS of the study areas and suggest a need for adequate management and crop protection during these critical crop development stages. Even with favourable conditions at these stages, grain yields rarely reach 5 Mg ha^{-1} for rice and sorghum. As mentioned earlier, the difference in terms of ‘Pest and disease impacts’, ‘Weed impacts,’ and ‘Water supply’ were considered over a relatively short period of time (mid stage of crop development only). Note also that crop yield was assessed in Mg ha^{-1} and a difference (e.g. 0.25 Mg ha^{-1}) that may be small in other areas is quite substantial for the farmers in our study areas.

Rice subjected to a high level of farming constraints during the initial and development stages seems insensitive to both water variation and changes in weed pressure (Fig. 10). However, controlling the level of pests and diseases, supported by good water supply, is likely to improve yield regardless of weed levels. Under limited water conditions and increasing levels of pests and diseases, however, weed removal seems to improve grain yield in rice. This slight yield improvement seems to level off with better farm conditions at the development stage. While weed removal generally improves rice grain yield, weeds are likely to have less of an effect when water is available, and pests and diseases are controlled. The effects of both weeds and pests and diseases seem to increase with increasing soil moisture.

In contrast to rice, sorghum grown under high farming constraints at the initial and development stages seems to respond to varying levels of water and weeds (Fig. 11). While the crop is likely to have higher yields with non-limiting water supply, it seems more suitable for drier conditions when severe pest and disease effects are expected. Only in this situation of strong pest and disease impact, however, does sorghum grown under non-limited water supply seem to show grain yield improvement due to weed reduction. Regardless of water conditions, improved farm management at the early stage seems to improve the effect of weed removal in sorghum, particularly when the level of pest and disease impact is minimal. With increased severity of pests and diseases, sorghum seems to no longer respond to weed reduction. When farm conditions are better at the development stage, both non-limiting water supply and weed reduction improve the yield of sorghum. When farm conditions are better at both initial and development stages, wetter conditions seem better under minimal pest and disease pressure, whereas drier conditions are better in situations of severe pest and disease pressure. In situations of limited pest and disease impact, sorghum seems to respond to weed reduction only when water is limited. Under severe pest and disease pressure, however, the crop seems to improve with weed removal regardless of the water conditions. This supports our earlier findings on the cumulative effect of farming constraints and

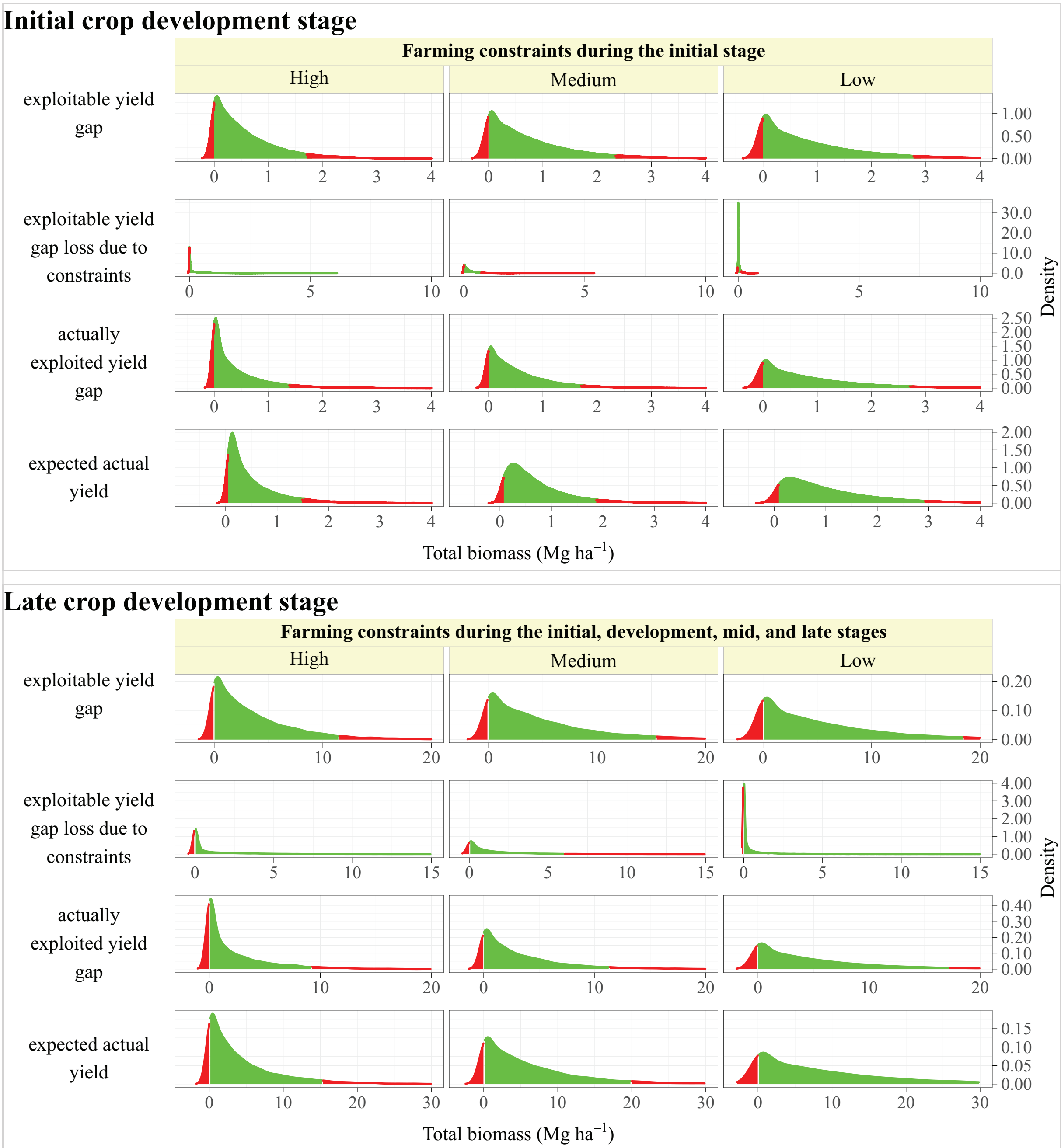

Fig. 8. Simulated biomass yield metrics for flood-based farming systems in the Tigray region of Ethiopia and Kisumu County in Kenya. The green and red colours, respectively, are values inside and outside the 90% confidence interval. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

suggests possible weed-sorghum competition for water.

4. Discussion

The model presented in this paper navigates system complexity by breaking down a complex agricultural system into meaningful components to provide a reasonable representation of the system. The participation of stakeholders and the inclusion of other relevant sources of information increases the chance that the model captures the necessary knowledge to adequately inform management decisions in FBFS (Krueger et al., 2012; Refsgaard et al., 2007). To provide decision support for development interventions, crop models generally need to include hard-to-measure aspects, which often require non-trivial measurement and analysis methods. For example, a crop model for FBFS cannot ignore the various social aspects, which are crucial for crop production. While the model development process can easily overlook such aspects, a crop model for agricultural decision support needs to consider the whole system and represent at least its most important factors. A model for such purposes should be framed in a way that allows the use of imperfect knowledge. New knowledge and high-quality data constitute a means of reducing initial uncertainties, but they should not be a prerequisite to running a model in the first place (Liman Harou et al., 2020).

The predictions of the model discussed in this paper have not been compared with actual yield data due to a lack of required secondary data in the study region and limited resources for collecting new data. Even if

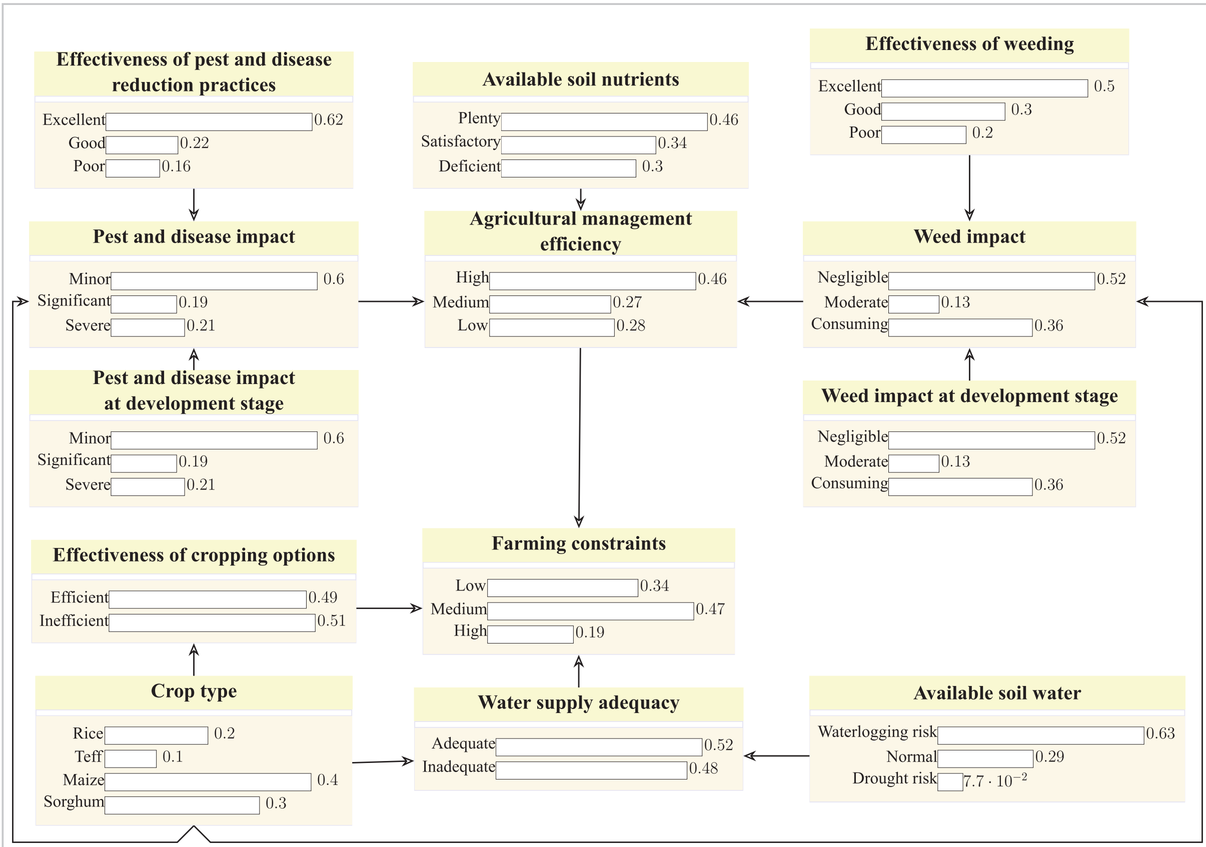

Fig. 9. Overview of causal relationships defining the farming constraints in flood-based farming systems in the Tigray region of Ethiopia and Kisumu County in Kenya.

more empirical measurements were available, they may still not meet the requirements for validating such probabilistic predictions that account for the entire space of possibilities. Still, by predicting all yield possibilities based on estimates from knowledgeable experts and on other relevant information available in the study regions, we have limited the risk of making unrealistic predictions. To support the use of local knowledge and improve model predictions, future studies may consider the use of biophysical information collected via field surveys and experiments (see Kilongosi et al., 2019), which was not possible in this study. The wealth of remote sensing data can be useful for both spatial and non-spatial models (Liman Harou et al., 2020). Available databases from previous modelling work, such as the global yield gap database, are potential sources of data for studies aiming to reproduce the proposed approach under rainfed conditions, but more caution is required for studies dealing with FBFS given the variation in FBFS settings. Even though the causal relationships between the variables are likely similar in many FBFS settings, studies aiming to use this model outside our study areas may need to specify the variable estimates according to specific local circumstances.

Soil type is crucial for water supply (Fig. 6), and soil properties are particularly important for the stability of water supply in FBFS settings. For example, clay soils are known to have high water holding capacity. Due to regular flooding, however, a soil's ability to store water can either decrease with the contribution of coarse sediments (i.e. gravel or sand) or increase with the contribution of finer sediments (i.e. silt and clay). Because coarse sediments tend to be deposited first, upstream soils are more likely to have low water retention than downstream soils.

A major concern for FBFS farmers is uncertainty surrounding the magnitude and intensity of flooding (Van Steenbergen et al., 2010). Water supply in FBFS does not only depend on the soil type or rainfall but also on the availability and management of agronomic flooding and soil fertility, along with the social organisation involved at various scales (Haile, 2010; Kilongosi et al., 2019; Van Steenbergen et al., 2010). Harmful consequences of water supply (e.g. drought, waterlogging) in a particular year can only be attenuated through adequate management. On soils with low water holding capacity (e.g. sandy soils), the use of groundwater, which is often relatively shallow due to the recurrent floods, can be considered for drought management. Groundwater from shallow wells can also constitute an additional source of drinking water, but little information is yet available regarding the quantity and quality of this water (Haile, 2010; Van Steenbergen et al., 2010). While soil fertility management can improve soil water conditions, the erosive effect of floodwater should not be underestimated. Rich soils are often transported between farmers' fields in the flood distribution network. The high risk of waterlogging on clayey and loamy soils can be addressed through limited tillage and enhanced soil organic matter content (Van Steenbergen et al., 2010).

The small effect of varying water levels and weeds on rice subjected to high farming constraints at the earlier development stages suggests that certain growth traits are affected by other forms of farming constraints than water and weeds, supporting the hypothesis that farming constraints during early growth are a major limitation to crop yield. Since with improved water supply, pest and disease control appears to result in higher yield, efforts to control pests and diseases would be more effective than weed removal, when conditions are harsh during the earlier stages of rice development. The slight yield improvement due to weed removal in situations of a high level of farming constraints suggests possible rice-weed competition for water, which becomes irrelevant with better farm conditions. The relatively high yield reduction due to pests and diseases in sorghum grown under non-limited water supply compared to dry conditions, and the yield improvement in this crop due to weed removal, suggest that healthy sorghum is likely to be weed-

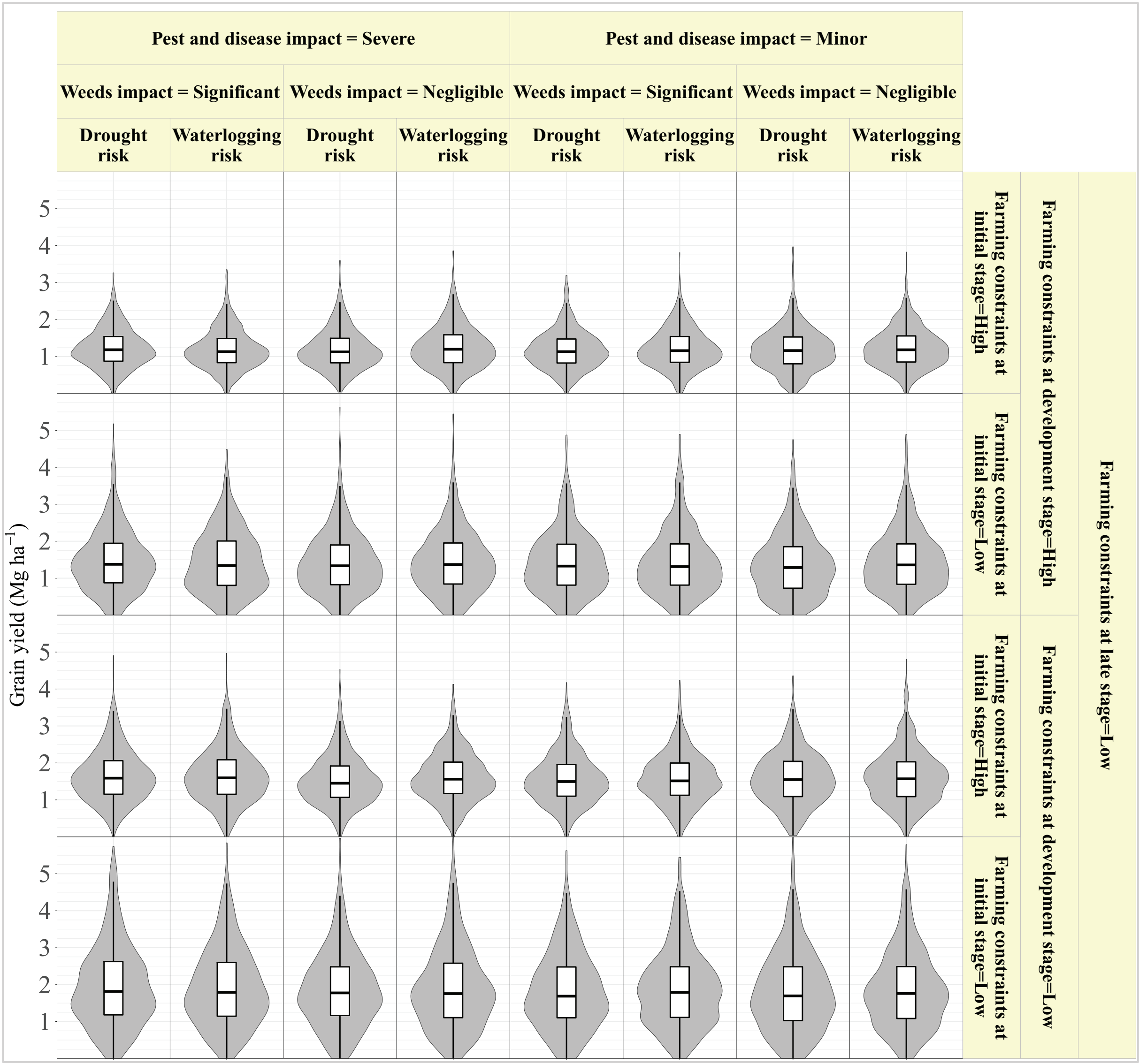

Fig. 10. Effects of pests and diseases, weeds and soil water on grain yield of rice grown in flood-based farming systems in the Tigray region of Ethiopia and Kisumu County in Kenya. Scenarios for water, weeds, pests and diseases are taken at the mid stage of crop development. Scenarios corresponding to 'Farming constraints at late stage=High' were not presented, because the results are barely visible owing to very small yield values below 1 Mg ha^{-1}.

tolerant to some extent, but weed removal may improve water productivity, particularly when weed-sorghum competition for space is limited. The lack of sorghum response to weed removal under severe pest and disease pressure implies the possibility of some sort of synergistic effect between these biotic stresses. Since sorghum responds better to improved water supply and weed reduction with better farm conditions at the development stage than rice, it is more likely than rice to escape from the farming constraint trap with improved management at later stages. The yield improvement in sorghum affected by pests and diseases due to drier conditions suggests that moisture exacerbates the effect of pests and diseases.

Despite the high potential for biomass production (Fig. 8), the net grain yield depends strongly on farming constraints at the late stage of crop development (mainly pests and diseases; Figs. 10 and 11), which are crucial in the FBFS of the study areas. The cumulative effect and variability of farming constraints across farmlands (Figs. 6, 8, 10 and 11) further illustrates the importance of recognizing the complex web of processes, many of which require further investigation. Future research hypotheses can be centred on the nexus between 4 important aspects:

1. Synergetic effects of different biotic stresses
2. Sensitivity of different crops to the joint effect of these stresses
3. Effect of water supply on different crops subjected to these stresses
4. Options for managing these stresses for particular crops

Another interesting venue for future studies would be the comparative analysis of the market values of different types of produce. This would require the use of time series data, since the market prices change throughout the year (Van Den Ham, 2008). When such aspects are well understood, experiments can be conducted to assess the attainable yield under reduced farming constraints and set realistic targets for closing the yield gap. According to Meng et al. (2013), three important factors are generally responsible for yield gaps. The first factor is a mismatch between the local climatic conditions and the crop cultivar, which does not allow exploitation of the entire available growing period. Inadequate sowing dates and thus relatively short-duration crops, can lead to earlier harvests than would be ideal, often not leaving enough time for crops to complete grain filling. In China, 7–15% of yield losses have been attributed to harvest before physiological maturity (Meng et al., 2013). Yield losses in FBFS are exacerbated by harvest and post-harvest losses

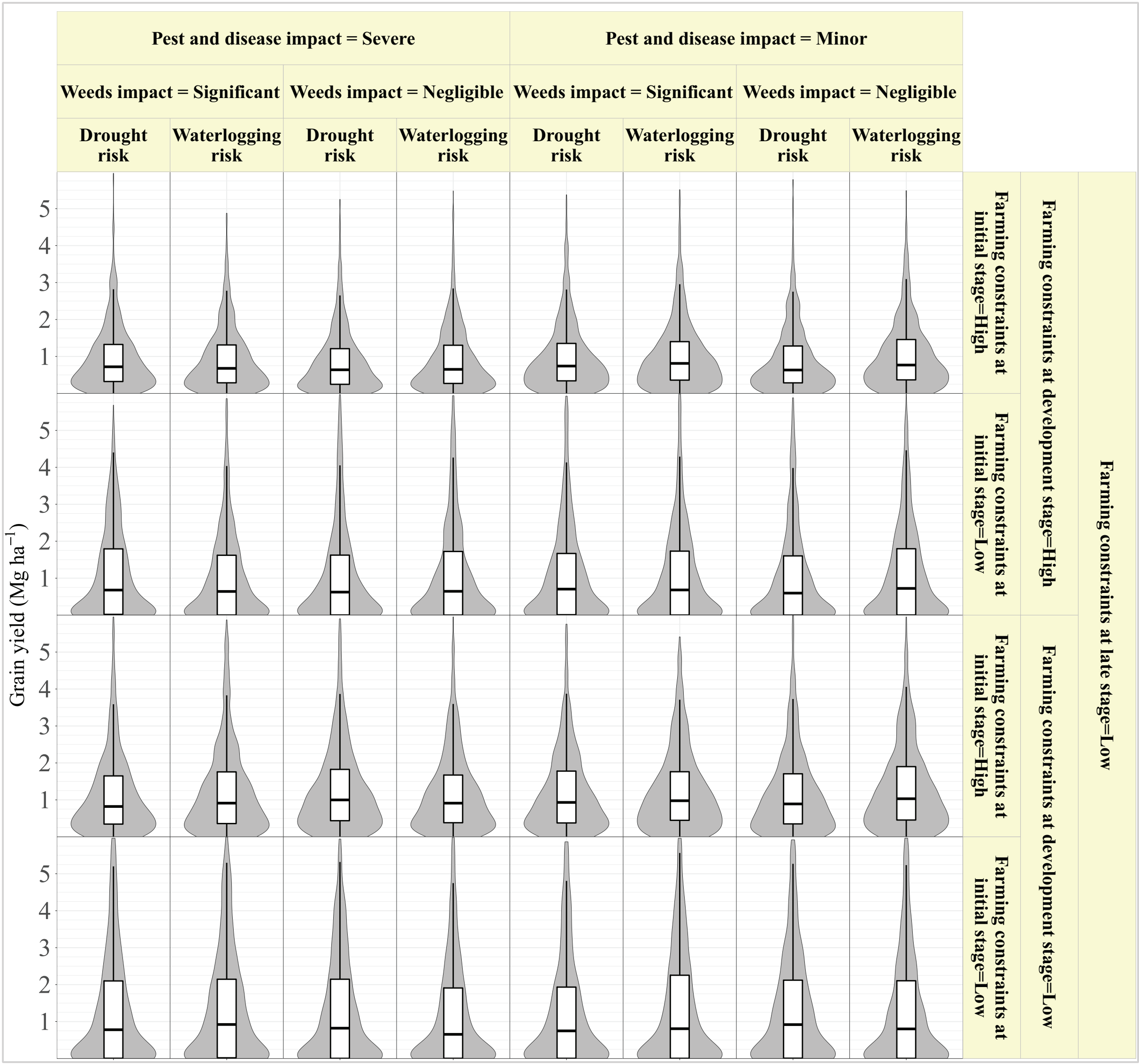

Fig. 11. Effects of pests and diseases, weeds and soil water on grain yield of sorghum grown in flood-based farming systems in the Tigray region of Ethiopia and Kisumu County in Kenya. Scenarios for water, weeds, pests and diseases are taken at the mid stage of crop development. Scenarios corresponding to 'Farming constraints at late stage=High' were not presented because the results are barely visible owing to very small yield values below 1 Mg ha^{-1}.

due to a range of pests and diseases (Figs. 10 and 11). In Kisumu for instance, rice farmers experience large losses due to bird attacks, besides other limitations of high agricultural importance. Other less perceptible losses, such as those incurred due to harvesting technologies, or market opportunity losses due to the lack of storage facilities, contribute to low returns on agricultural investments (Van den Berg and Singels, 2013). The second factor responsible for yield gaps is related to the inadequacy of input allocation (e.g. inadequate water and nutrient supply). In Tigray, and to some extent also in Kisumu, few farmers use recommended amounts of fertilizers, while many schemes are exposed to water shortages. The third major reason for yield decline is poor management resulting in inadequate crop protection.

5. Conclusions and policy recommendations

This paper presents a novel approach to crop modelling that addresses system complexity in the context of limited information. Using Decision Theory methods, we demonstrate how expert knowledge can be used to support decision making under uncertainty. Using the example of FBFS, we build a complex system model using simple building blocks and imperfect information and demonstrate how to make predictions based on different scenarios that account for variability across farmlands within a farming system. We show how intangible factors can be assessed for inclusion into the model and how seemingly marginally important factors can have cumulative effects on model predictions. The findings suggest that addressing the major farming constraints, such as uncertainty in water supply or the production risk due to biotic stresses, could reduce yield gaps and provide as much as double the current crop production in FBFS of the study areas. Therefore, development policies should focus on sustainable water supply and crop protection. Research and development should concentrate on closing the current FBFS yield gap through the study of farming constraints that compromise primary production. From a modelling perspective, more robust assessments of agricultural systems are possible using empirical models, preferably derived from field experiments, in a probabilistic framework. Other promising avenues could be the modelling of agricultural systems with consideration of system variability and model boundary conditions in a spatially explicit fashion. We hope the approach we presented will stimulate robust assessments that adequately account for system complexity and data uncertainty.

Such predictions can be instrumental for enhancing the usefulness of model-based decision support in complex real-world settings.

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.

Acknowledgements

This study was funded by the Deutscher Akademischer Austauschdienst (DAAD) and the World Agroforestry Centre (ICRAF). We thank the farmers and experts who contributed to the conceptualization and development of the model. We thank the department of Dryland Agriculture of Mekelle University for guidance on the sampling frame in the Tigray region and for facilitating logistics during field work. We are also grateful to the participants of the leadership course in flood-based farming and water harvesting in Kenya and the participants of the International Training on Integrated Watershed Management and FBFS in Ethiopia, who helped in drafting the models.

Software and data availability

We conducted data analysis using the R programming language (R Core Team, 2019). We provided two GitHub repositories with replicable data and code. The core code, available in one repository (https://github.com/Issoufou-Liman/decisionSupportExtra), is an R package describing the major functions developed during the preparation of the manuscript. The data along with additional code are available in another repository (see https://github.com/Issoufou-Liman/Modelling_FBFS/releases/tag/0.0.0.9000).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.agsy.2020.103014.

References

Allen, R.G., Pereira, L.S., Raes, D., Smith, M., 1998. Crop Evapotranspiration: Guidelines for Computing Crop Requirements, FAO Irrigation and Drainage Papers (Book 56). FAO, Rome, Italy.

Baroni, G., Tarantola, S., 2014. A general probabilistic framework for uncertainty and global sensitivity analysis of deterministic models: a hydrological case study. Environ. Model. Softw. 51, 26–34. https://doi.org/10.1016/j.envsoft.2013.09.022.

Belgorodski, N., Greiner, M., Tolksdorf, K., Schueller, K., 2017. rriskDistributions: Fitting Distributions to Given Data or Known Quantiles.

Boote, K.J., Jones, J.W., Pickering, N.B., 1996. Potential uses and limitations of crop models. Agron. J. 88, 704–716. https://doi.org/10.2134/agronj1996.00021962008800050005x.

Delignette-Muller, M.L., Dutang, C., 2015. fitdistrplus: an R package for fitting distributions. J. Stat. Softw. 64, 1–34. https://doi.org/10.18637/jss.v064.i04.

Embaye, T.G., Beevers, L., Mehari Haile, A., 2012. Dealing with sedimentation issues in spate irrigation systems. Irrig. Drain. 61, 220–230. https://doi.org/10.1002/ird.630.

Erkossa, T., Langan, S.J., Hagos, F., 2014. Constraints to the development, operation and maintenance of spate irrigation schemes in Ethiopia. In: Erkossa, Teklu, Hagos, Fitsum, Lefore, N. (Eds.), Proceedings of the Workshop on Flood-Based Farming for Food Security and Adaption to Climate Change in Ethiopia: Potential and Challenges. International Water Management Institute (IWMI), Adama, Ethiopia, pp. 5–22. https://doi.org/10.5337/2014.233.

FBLN, 2018. Flood-Based Livelihood Network (FBLN) Foundation [WWW Document]. URL. http://spate-irrigation.org/.

Fenton, N., Neil, M., 2013. Risk Assessment and Decision Analysis with Bayesian Networks. CRC Press, Boca Raton, Floria, USA.

Gallagher, P., 1987. U.S. soybean yields: estimation and forecasting with nonsymmetric disturbances. Am. J. Agric. Econ. 69, 796–803. https://doi.org/10.2307/1242190.

Haile, A.M., 2010. A Tradition in Transition, Water Management Reforms and Indigenous Spate Irrigation Systems in Eritrea, 1st ed. CRC Press, London.

Hansson, F., Sjökvist, S., 2003. Modelling Expert Judgement into a Bayesian Belief Network – A Method for Consistent and Robust Determination of Conditional Probability Tables. LUND University.

Harlan, J.R., Pasquereau, J., 1969. Décrue agriculture in Mali. Econ. Bot. 23, 70–74. https://doi.org/10.1007/BF02862973.

Højsgaard, S., 2012. Graphical Independence Networks with the gRain Package for R. J. Stat. Softw. 46, 1–26. https://doi.org/10.18637/jss.v046.i10.

Hubbard, D.W., 2014a. How to Measure Anything: Finding the Value of Intangibles in Business, 3rd ed. John Wiley & Sons, Inc., Hoboken, New Jersey, USA.

Hubbard, D.W., 2014b. How to Measure Anything Third Edition Calibration Questions [WWW Document]. URL. https://www.howtomeasureanything.com/wp-content/uploads/2014/02/HTMA_3rd.ed_Calibration_Questions.doc. accessed 10.9.20.

Jax, K., 2005. Function and "functioning" in ecology: what does it mean? Oikos. https://doi.org/10.1111/j.1600-0706.2005.13851.x.

Jensen, F.V., 1996. An Introduction of Bayesian Network, 1st ed. Springer, Har/Dskt edition, Heidelberg.

Junk, W., Bayley, P., Sparks, R., 1989. The flood pulse concept in river – floodplain systems. In: Dodge, D.P. (Ed.), Proceedings of the International Large River Symposium. Can. Spec. Publ, Honey Harbour, Ontario, Canada, pp. 110–127.

Kilongosi, C., Raude, J., Wambua, R., Malesu, M., Oduor, A., 2019. Assessment of the best moisture conservation technique in Sorghum production under spate irrigation: a case study of Ewaso Nyiro south Drainage Basin. J. Sustain. Res. Eng. 5, 12–24.

Kitagawa, G., 1996. Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. J. Comput. Graph. Stat. 5, 1–25. https://doi.org/10.2307/1390750.

Koller, D., Friedman, N., 2009. Probabilistic Graphical Models: Principles and Techniques, 1st ed. MIT Press, Cambridge, Massachusetts, USA.

Krueger, T., Page, T., Hubacek, K., Smith, L., Hiscock, K., 2012. The role of expert opinion in environmental modelling. Environ. Model. Softw. 36, 4–18. https://doi.org/10.1016/j.envsoft.2012.01.011.

Liman Harou, I., Whitney, C., Kung'u, J., Luedeling, E., 2020. Mapping flood-based farming systems with Bayesian networks. Land 9. https://doi.org/10.3390/land9100369, 369.

Luedeling, E., Shepherd, K., 2016. Decision-focused agricultural research. Solutions 7, 46–54.

Luedeling, E., Oord, A.L., Kiteme, B., Ogalleh, S., Malesu, M., Shepherd, K.D., De Leeuw, J., 2015. Fresh groundwater for Wajir – ex-ante assessment of uncertain benefits for multiple stakeholders in a water supply project in Northern Kenya. Front. Environ. Sci. 3 https://doi.org/10.3389/fenvs.2015.00016.

Luedeling, E., Goehring, L., Schiffers, K., 2019. decisionSupport: Quantitative Support of Decision Making under Uncertainty (R package version).

Meng, Q., Hou, P., Wu, L., Chen, X., Cui, Z., Zhang, F., 2013. Understanding production potentials and yield gaps in intensive maize production in China. Field Crop. Res. 143, 91–97. https://doi.org/10.1016/j.fcr.2012.09.023.

Murthy, V.R.K., 2004. Crop growth modeling and its applications in agricultural meteorology. In: Sivakumar, M.V.K., Roy, P.S., Harmsen, K., Saha, S.K. (Eds.), Satellite Remote Sensing and GIS Applications in Agricultural Meteorology. World Meteorological Organisation, Dehra Dun, India, pp. 235–261.

Pearl, J., 1988. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

Puertas, D.G.-L., van Steenbergen, F., Haile, A.M., Kool, M., Embaye, T.G., 2011. Flood Based Farming Systems in Africa (No. 5), Overview Paper. The netherland.

R Core Team, 2019. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Ramirez, O.A., Misra, S., Field, J., 2003. Crop-yield distributions revisited. Am. J. Agric. Econ. 85, 108–120. https://doi.org/10.1111/1467-8276.00106.

Refsgaard, J.C., van der Sluijs, J.P., Højberg, A.L., Vanrolleghem, P.A., 2007. Uncertainty in the environmental modelling process – a framework and guidance. Environ. Model. Softw. 22, 1543–1556. https://doi.org/10.1016/j.envsoft.2007.02.004.

Rosenstock, T.S., Mpanda, M., Rioux, J., Aynekulu, E., Kimaro, A.A., Neufeldt, H., Shepherd, K.D., Luedeling, E., 2014. Targeting conservation agriculture in the context of livelihoods and landscapes. Agric. Ecosyst. Environ. 187, 47–51. https://doi.org/10.1016/j.agee.2013.11.011.

Scutari, M., 2010. Learning Bayesian Networks with the bnlearn R Package. J. Stat. Softw. 35, 1–22. https://doi.org/10.18637/jss.v035.i03.

Scutari, M., Denis, J.-B., 2015. Networks Bayesian with Examples in R, 1st ed. CRC Press, Boca Raton.

Uusitalo, L., Lehikoinen, A., Helle, I., Myrberg, K., 2015. An overview of methods to evaluate uncertainty of deterministic models in decision support. Environ. Model. Softw. 63, 24–31. https://doi.org/10.1016/j.envsoft.2014.09.017.

Van den Berg, M., Singels, A., 2013. Modelling and monitoring for strategic yield gap diagnosis in the south African sugar belt. Field Crop. Res. 143, 143–150. https://doi.org/10.1016/j.fcr.2012.10.009.

Van Den Ham, J.-P., 2008. Dodota Spate Irrigation System Ethiopia – A Case Study of Spate Irrigation Management and Livelihood Options. Wageningen University. https://edepot.wur.nl/14997.

Van Ittersum, M.K., Cassman, K.G., Grassini, P., Wolf, J., Tittonell, P., Hochman, Z., 2013. Yield gap analysis with local to global relevance-a review. Field Crop. Res. 143, 4–17. https://doi.org/10.1016/j.fcr.2012.09.009.

Van Steenbergen, F., Lawrence, P., Haile, A.M., Salman, M., Faurès, J.-M., 2010. Guidelines on Spate Irrigation. FAO, Rome, Italy.

Van Steenbergen, F., MacAnderson, I., Mehari, A., 2011. Spate Irrigation in the Horn of Africa: Status and Potential. (No. 2), Overview Paper. The Netherlands.

Whitney, C., Lanzanova, D., Muchiri, C., Shepherd, K.D., Rosenstock, T.S., Krawinkel, M., Tabuti, J.R.S., Luedeling, E., 2018a. Probabilistic decision tools for determining impacts of agricultural development policy on household nutrition. Earth's Futur. 6, 359–372. https://doi.org/10.1002/2017EF000765.

Whitney, C., Shepherd, K., Luedeling, E., 2018b. Decision Analysis Methods Guide; Agricultural Policy for Nutrition (No. Working Paper No. 275). Nairobi. https://doi.org/10.5716/WP18001.PDF.

Liman Harou, I., Whitney C., Kung'u, J., Luedeling, E., 2020. Modelling Risk and Uncertainty in Flood-based Farming Systems. https://github.com/Issoufou-Liman/Modelling_FBFS/releases/tag/0.0.0.9000.GitHub. accessed 16.12.20.