

Modelling Risk and Uncertainty in Flood-based farming systems.

Issoufou Liman^{1,2,}*

Cory Whitney^{1,3}

James Kungu²

Eike Luedeling^{1,3,4}

May 07, 2019

Abstract

Modelling the performance of crops grown under specific conditions is an important tool for decision support. Towards this end, many crop models are being used to inform agricultural policies. However, many of these models are applied beyond a reasonable scope of application, mainly because they are either designed for other settings, they fail to capture system complexity, or because some of their data requirements are not within reach. In this paper, we demonstrate an approach to overcome the above-mentioned problems using a probabilistic crop model based on the Decision Theory methods. Using the innovative method, we transparently use all available sources of information and consider various sources of uncertainty using participatory analysis. The transdisciplinary process provides a model where all important variables and their interactions, as determined by local experts and relevant literature, are used to simulate the plausible ranges of expected grain and biomass yield at various stages of crop development. System components are described using individual Bayesian networks in which expert knowledge is employed to specify causality among important variables along with their probability distributions. These Bayesian networks, along with other quantitative variables, are used as input for a Monte Carlo model to simulate the plausible performance of crops. Our objective is to develop a customizable solution-oriented approach for crop modelling in settings where data and resources are limited. We describe how such a model can be developed and provide three case study applications focusing on flood-based farming systems. The case studies illustrate the extensible scope of such a model by assessing the performance of a cropping system, individual crops, or other factors of high agricultural importance across the Tigray region of Ethiopia and Kisumu County of Kenya. The probabilistic tools and transdisciplinary approaches of Decision Theory allow experts and analysts to address the system complexity and data uncertainties in crop modelling.

Contents

I. Introduction	3
II. Study Area	4
IIIMaterials and methods	6
III.1. Conceptual Framework	6
III.1.1. Concepts and Rational	6
III.1.2. Methodological Specifications	7
III.2. Data acquisition	8

III.3. Data processing	9
III.3.1. Making conditional probability tables (CPTs)	9
III.3.2. The models and the modelling interface	10
IV.Result	10
IV.1. Overview of the Conceptual Model	10
IV.2. Case studies	11
IV.2.1. Qualitative Soil water Assessment under different Management Decisions at initial stage of crop development in Flood-based livelihood systems.	11
Abstract	11
Introduction	11
Materials and Method	12
Results and Discussions	12
IV.2.2. Probabilistic Assessment of Biomass Accumulation in Major Crops grown under FBFS in Kisumu, Kenya and Tigray, Ethiopia.	13
Introduction	13
Materials and Method	13
Results and Discussions	13
IV.2.3. Quantitative grain yield variability and relative performance of Rice and Sorghum under different scenarios for water and crop enemies at mid-stage of crop development in flood-based Agriculture.	14
Introduction	14
Materials and Methods	14
Results	14
Main effects of the different treatments on grain yield	15
Marginal and Joint effects of the levels of the different treatments on grain yield	15
V. Discussion	16
reference	16

- ¹ World Agroforestry Centre (ICRAF), United Nations Avenue, Gigiri, P. O. Box 30677-00100, Nairobi, Kenya
- ² Kenyatta University, Department of Environmental Sciences
- ³ Center for Development research (ZEF), University of Bonn, Genscherallee 3, D-53113, Bonn, Germany
- ⁴ University of Bonn, Department of Horticultural Sciences, Auf dem Hügel 6, D-53121, Bonn, Germany

* Correspondence: Issoufou Liman <issoufoul@gmail.com>

I. Introduction

Modelling the performance of crops grown under specific conditions provides useful support for decision making. There are many models to simulate agricultural production.¹ These models have proven useful for various science and policy applications.^{2,3} While these models are quite sophisticated, a few complications limit their broad applicability. One important pitfalls of many models is their failure to describe reality sufficiently well to support real-life decisions.^{3,4} Consequently, they fail to address their dual purpose of generating knowledge while contributing actionable information to support decisions and ultimately generate impacts.⁴ These limitations of traditional modelling approaches can be explained by many issues among which their inadequacy to account for systems' complexity, the scope of application, the available data, or customization in specific settings.^{1-3,5-7} Contrastingly, the development of new models that address these challenges while producing reliable results for solving real-world problems has proven to be difficult in many contexts. In crop modelling, these problems often result in unrealistic yield assessments,^{1,7} which highlight the need for better approaches that accommodate for systems' complexity and imperfect information. We argue that crop models should consider data uncertainty and find ways to include data inputs of varying nature, origin and quality. We demonstrate the usefulness and usability of this concept by developing a flexible crop model for Flood-based Farming Systems (FBFS). FBFS are agricultural systems receiving additional irrigation from various types of non-harmful floods that are unpredictable, of short duration, of low annual frequency and occurring in relatively low-lying areas with gentle topography. In FBFS, water supply to crops is often determined by complex socio-institutional arrangements for water acquisition and sharing.^{8,9} In this contexts, floods are not natural disasters but temporary situations of excess water, driven by the natural variability as in the concept of flood pulse,¹⁰ or in the concept of Crue/Décrue¹¹ such that the water availability is governed by recurrent floods followed by recession providing opportunities for farming. Many authors¹¹⁻¹⁴ have stressed their importance for rural livelihoods in different communities. Water management practices in FBFS take advantage of extreme rainfall variability by saving water against potential drought. However, FBFS also pose many challenges for agricultural production,^{9,13} many of which are related to the timing, frequency or size of flood.¹³ These risks must be appreciated in a context of a general lack of information, which makes evidence-based management very difficult.⁹ The lack of information on FBFS may be due to a general lack of attention to FBFS in agricultural research and policy,^{9,13} which has limited the build-up of expertise and comprehensive information on what management practices are best suited to FBFS.¹⁴ The flood-related risks and uncertainties in information and agricultural production are too tangible to disregard. Consequently, farmers and donors are reluctant to invest in FBFS despite their productivity because of the lack of risk assessment procedures suited to FBFS. Such a decision support gap results in substantial opportunity loss in many context.^{15,16} The promises of spate irrigation (a variant of FBFS) compared to rainfed agriculture has been reported in Kobo (Ethiopia) as four times and double yield increase respectively for pepper and sorghum.¹⁴ Such positive stories will, however, only materialize when the additional water is well distributed in time and complements the shortages in rainwater. Furthermore, floods may also result in excessive water, which may cause waterlogging of soils or submergence of crop canopies with negative effects on crop production. FBFS are clearly promising and complex socio-ecological systems but also prone to many risk factors and data uncertainty. In FBFS, the above-mentioned limitations of traditional modelling approaches are further exacerbated by socio-economic and management aspects that are crucial for system functioning. FBFS Models must consider sediment management, infrastructure for water acquisition, social rules and other institutional arrangements for water acquisition and sharing among farmers. While these aspects are yet to be addressed in crop models as mentioned earlier, there is so far no crop model for FBFS.

in our knowledge. The objective of this paper is to demonstrate how to develop a crop model for complex agricultural systems with limited information while accounting for data uncertainty. We applied decision theories methods to the concept of FBFS considering the available information and factors that seem important to ultimately simulate the performance of some crops along with their expected production risks. For the model to understand the farmers ‘realities, it needs to be able to account for qualitative as well as quantitative information. We therefore used Bayesian Networks (BNs) within a framework of Monte Carlo (MC) models to develop a crop model. The BNs were used to describe important qualitative processes, mostly related to agricultural constraints (e.g. the cropping systems and management options adopted by the farmer), and the MC models were used to describe the quantitative ones (e.g. biomass accumulation across crop development stages). FBFS are considered as showcase but the concept is customizable to any other complex systems particularly 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 contexts with minimal modification. The two areas are assumed to provide most of the complexity (e.g. bio-physical systems type, agricultural managements, social and institutional arrangements) found in FBFS.

II. Study Area

The study areas, namely the Kisumu County and the Tigray region respectively in Kenya and Ethiopia, are located within relatively low-lying lands with differences regarding the management of agronomic flooding and other agricultural practices (Figure 1). While FBFS in Kisumu mainly source their water from permanent reservoirs via inundation canals, those in Tigray mostly obtain their water from ephemeral rivers where farmers are required to divert important amounts of spate flow within a relatively short period of time. Communities in the two study areas employ specific strategies for flood and land management (Figure 1). These differences were addressed by a stratified sampling frame applied during field work. In the Tigray region (Ethiopia), three schemes including Tsige’ a (39.59967° N, 12.79827° E), Harosha (39.62788° N, 12.31102° E), and Dayu (39.60545° N, 12.47372° E) were sampled to capture different flood water diversion strategies. While the flood is diverted using improved water diversion in Tsige’ a, traditional and modern ones are found respectively in Harosha and Dayu. Here, traditional flood water diversions are to be understood as physical infrastructures (e.g. deflecting spurs, soil bunds) constructed by the farmers across flood channels using locally available materials. In modern ones, concrete infrastructures (e.g. diversion weirs) designed by engineers replace those in traditional diversion. The improved diversion type consisted in integrating both farmers and engineer knowledge in designing the infrastructures. In Kisumu county (Kenya), where five areas (Awach Kano (0.22762° S, 34.95432° E), East Kano (0.14700° S, 34.93436° E), West Kano (0.18358° S, 34.81306° E), Out-growers (0.18822° S, 34.89974° E), East Nyankach (0.27578° S, 34.97058° E)) were considered, different situations are encountered. In Awach Kano, water is acquired via simple gravity flow from the Awach river, whereas in East Kano and West Kano it is acquired by large pumps from Nyando River and Lake Victoria, respectively. It worth mentioning that the East and West Kano schemes are conventional irrigation schemes managed by the Kenyan National Irrigation Board (NIB), even though they share many properties with FBFS. The scheme of the so called Out-growers is also fed via gravity as in the case of Awach Kano from the river Nyando, but it differs from the other locations in that it also serves as a safe disposal for excessive water in East Kano. In East Nyankach, water is obtained through runoff harvesting via various household ponds and water tanks. These differences have important implications for water supply, social organization, and crops grown under FBFS (Figure 1). For example, risks related to water supply may be more prominent in the Out-growers’ scheme

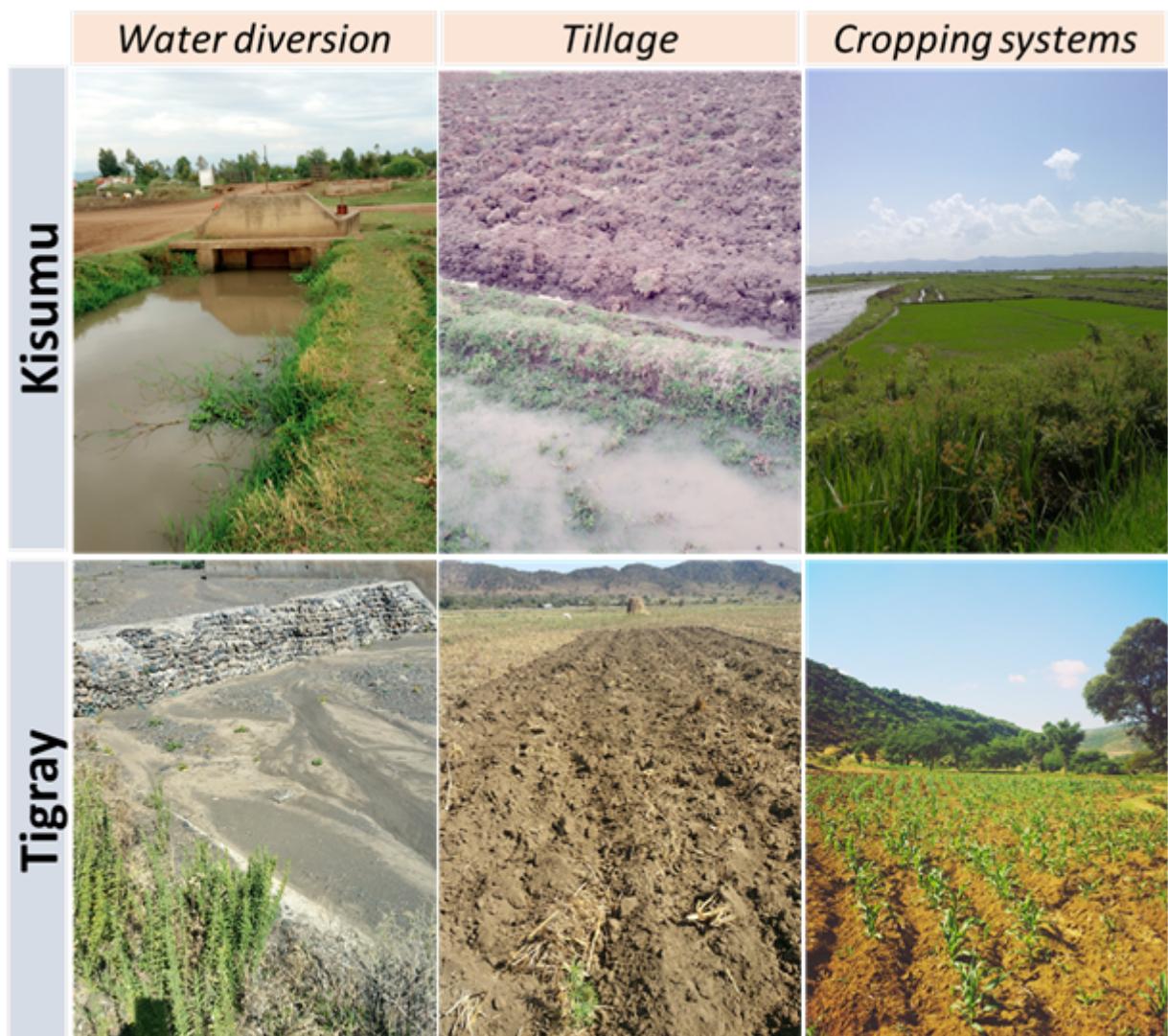


Figure 1: Differences in the practices of FBFS in Tigray region (Ethiopia) and Kisumu County (Kenya).

compared to the schemes managed by the NIB. While the water sources and acquisition procedures vary under rice-based systems in Kisumu (i.e. Awach Kano, East Kano, West Kano, Out-growers), cultural practices are similar. In these areas, rice is sown (after tillage and flooding) in monocultures during floods, followed by various types of flood recession crops that are intercropped depending on the farmer and water availability (Figure 1). In Tigray, however, crops (mainly Maize, Sorghum, Teff) are mostly sown based on rainfall to later be irrigated using flood water (Figure 1).

III. Materials and methods

III.1. Conceptual Framework

III.1.1. Concepts and Rational

The modelling concept is based on two different probabilistic approaches, the MC simulation for handling continuous variables, and the BNs for handling the discrete ones, as mentioned earlier. The resulting models, embedding complex relationships among several processes of high agricultural importance, were interfaced (Figure 2) to provide quantitative assessment in FBFS. Our goal was to develop a detailed description of the important processes, highlighting the factors of risk and uncertainty, that influence crops development and production focusing on various yield metrics (Figure 2). Towards this end, the conceptual framework (Figure 2) defines a farmland as the smallest entity evolving within a complex farming system where processes, fulfilling specific functions, are driven by several variables. These variables defining various biotic and abiotic factors are, in turn, moderated by a package of agricultural practices.¹⁷ The interactions between these and the resulting low-level functions and processes profile the farmland in terms of relevant farming constraints. If the processes leading to farming constraints vary according to farmland, the variable farming constraints also varies, then they can be assessed in terms of their relative effects on crops at farm levels. Consequently, the variable farming constraint can be described using the scored values of these effects as probability distribution when these effects are measurable at plot levels, hence across the higher order farming system. Since the variables involved in the processes are causally related, the system complexity can be unambiguously dealt with by understanding the individual processes (e.g. (causality chains leading to the) available soil water or available soil nutrients) within the system. These processes, in turn, can be decomposed into the respective variables (e.g. rainfall occurrence or manure application) forming them. Furthermore, the variability in the processes across farmlands implies further state changes in the variables forming them, in which case each variable can be defined in terms of the different states (i.e. rainfall occurrence is either on-site or off-site, manure application is either true or false) it may take at farming plot levels. Such complex levels of hierarchy and categorization are the domain of multinomial models and can be well specified using BNs whereas the spatio-temporal behaviour of the system can be quantitatively assessed using MC models considering plausible sub-samples across the space of the probability distribution. With this assumption, causality chains leading to key intermediate variables were used to ultimately assess the level of farming constraints using low-level BNs estimating the probability of farming constraint at plot level. These local BNs were connected to be used as synthetic inputs for the MC model (Figure 3; Suppl.) describing the crop development stages following the FAO nomenclature.¹⁸

A Monte Carlo simulation is an intensive computation, where each unknown variable is estimated using mathematically valid rules, resulting in different scenario depending on the probabilities of the input data. These redundant calculations are repeated thousands of time to generate the scenario to portray the range of possible outcomes along with their probabilities. The inputs for MC models

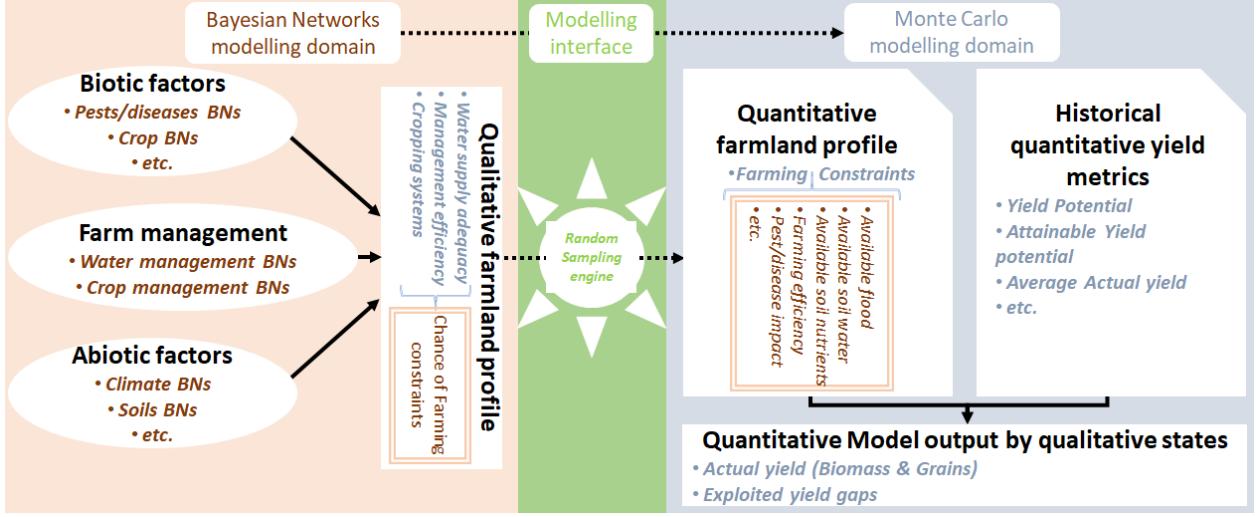


Figure 2: Conceptual framework of important processes to include in the development of a mixed model describing FBFS in Kenya and Ethiopia.

are continuous variables provided as probability distributions bounded by realistic value ranges to represent the associated uncertainty. BNs are probabilistic graphical model representing logical connections (directed edges) between a set of variables (nodes) where dependencies are represented by arrows (arcs) encoding the direction, and nature (i.e. direct or indirect relationship) of the causality.¹⁹ A Node describes the states a variable (node states) whereas a variable at the arrow head (child node) is influenced by the variable at the arrow tail (parent node). The strength of the influence is portrayed by the conditional probabilities.^{19–21} The properties described above make the MC models suitable for quantitative assessment,²² BNs suitable for predicting the extent to which specific effects are produced by certain causes,^{19–21,23} and both MC models and BNs suitable for imperfect information, expert knowledge, and explicit uncertainty. While MC models provide a means for quasi-exhaustive enumeration across the space of possibilities for quantitative risk assessment under uncertainty,²⁴ BNs are suitable for building discrete models describing different scenarios for qualitative variables.^{19,21} We represent discrete nodes as instances of a finite set of mutually exclusive states (node states) along with their occurrence probabilities²¹ as conditional probability tables (CPTs; see also section III.3.1) where the states of a parent node determines that of the child node.^{20,23,25} The MC model, in turn, was used for approximating probability distributions of unknown variables using probabilistic sampling by generating multiple random realisations of numerical equations aiming at quantitative biomass and grain yield assessment.^{22,26}

III.1.2. Methodological Specifications

We use an MC model to specify deterministic relationships targeting several unknown (dependent) variables (Figure 2) and estimate them based on random values of the known (independent) ones used as inputs for deterministic relationship linking the dependent to the independent variables.^{24,26} The BNs were used as multivariate models specifying the probabilities of the states of unknown variables conditional on the probabilities of the states of the known ones (Figure 2) as specified in a causal relationships between these discrete variables.^{19,21} The MC model was based on the ‘yield gap’ concept where yield is crop-specific, primarily determined by the genetic potential of the variety, the cultivar, and the farming constraints. We assume the yield potential to be the highest

possible theoretical yield that is never achieved due to some prevailing constraints that are context-specific⁷. Since these constraints vary at various spatial scales (e.g. at watershed level due to natural variability, at farmland scale due to management variability), losses in yield potential will also vary. Such losses can be estimated via the probability of the states of these local constraints. Then, the actual yield at a given plot is proportional to the probability of the state of local constraints at that plot. Four crops (rice, maize, sorghum and teff) were chosen based on their importance in the study areas. A high-level overview of the model is provided in section IV.1 and a more detailed description in the supplementary materials. Crop development stages were defined from the onset 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.¹⁸ At each of these 4 stages, local BNs were used to describe the farming plot and assess the probability of farming constraints which is assumed to inform the quality of management practices employed by the farmer. The farming constraints variable, in turns, is assumed to drive several yield metrics tracing the temporal biomass accumulation and potential opportunity losses with regards to yield improvement. The yield metrics include the yield potential defined as the cultivar yield without farming constraints (i.e. limitations in water, nutrient, and stress), the exploitable yield potential defined as the fraction of yield potential (70 – 80%) achievable by the farmers, the attainable yield potential defined as the maximum achievable yield given the farmers' realities, the average actual yield defined as the realistic yield actually achieved by the farmer, the exploitable yield gap defined as the yield improvement opportunities for the farmer, the exploitable yield gap loss due to farming constraints defined as the yield improvement opportunity loss, the actual exploited yield gap defined as the real yield improvement actually achieved by the farmer, and the expected actual yield defined as yield expectation given the farmer's management conditions.⁷ Note, some of these metrics at the next crop development stage depend on those at the previous stage whereas a biomass expansion factor, measuring the growth rate of the crop, was considered to control the crop growth at each of these stages. This way, crops would not arbitrary grow beyond the actual yield expectations. This cross-stage interdependencies holds also for some of the variables in BNs to account for residual effects of certain variables.

III.2. Data acquisition

The mixed BN / MC model was structured to account for causality deemed important to FBFS settings by considering various sources of information, including literature, online databases, expert knowledge elicitation, farmer and expert interviews and remote sensing. The model development process consisted of 5 sequential steps aiming at developing theories to be contextualised with farmers' realities (Figure 3). We reviewed the essential literature regarding the subject (literature review) to understand the concept, design broad leading questions, and target potential experts for interviews (high level discussion with experts). These high-level discussions provided the theoretical framework and leading questions for focus group discussions with farmers and meetings with local experts upon which individuals farmers interviews were conducted (Figure 3). The literature review covered most of the flood-based livelihood network database¹² and extended to other relevant sources. The high-level discussion involved 11 academic subject matter experts working with FBFS and related fields across the world. The focus group discussions, the local expert meetings and the farmer interviews were conducted in Ahero and Kisumu towns (Kenya) in December 2016 and June 2017 and Mekelle and Alamata towns (Ethiopia) in December 2016 and January 2017. Around 20 focus group discussions and 159 farmer interviews were conducted. The pool of local experts, initially represented by local farmers and extension civil servants, was extended to include 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.¹² Prior to the model development exercise, the experts were taken through the principles of decision analysis, calibration training^{22,26–28} and the available data. Following the 5 steps, we adopted a procedure where participants led the discussions depending on their expertise.

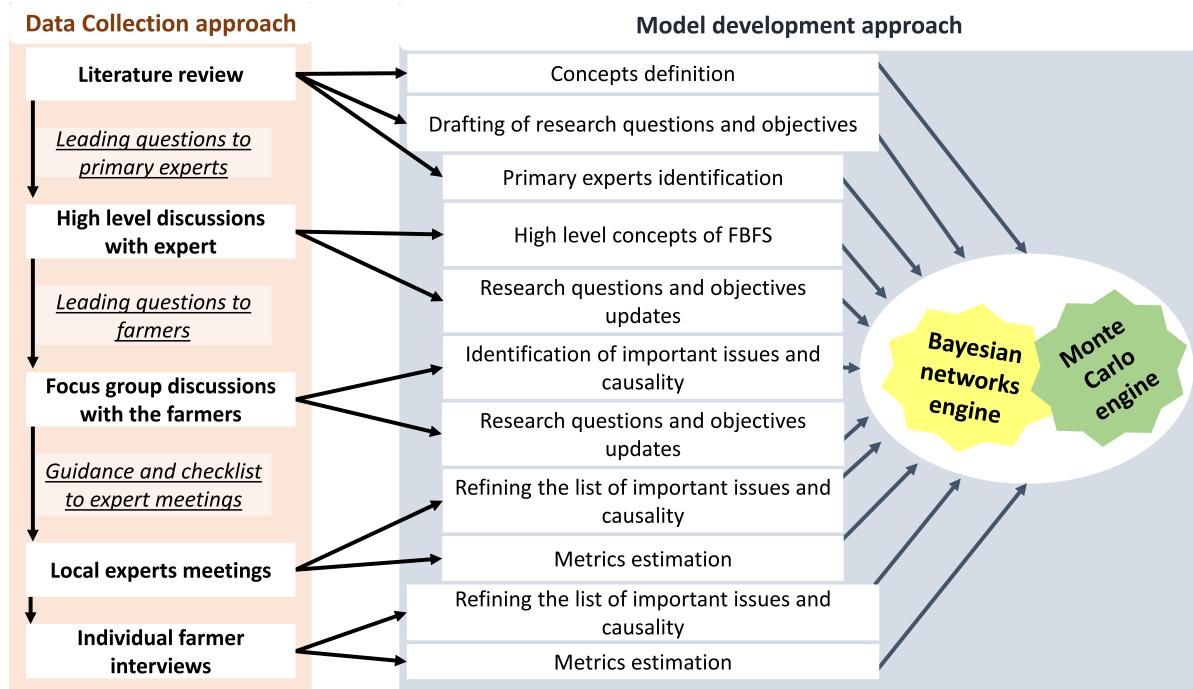


Figure 3: Overview of the approach used to develop the crop model for FBFS in Kenya and Ethiopia.

III.3. Data processing

III.3.1. Making conditional probability tables (CPTs)

As mentioned in section III.1, BNs are complex model in which influence between child and parent nodes are represented by arcs whereas the strength of that influence is represented by the conditional probabilities. Conditional probability tables (CPTs) are stochastic matrix storing these probabilities to portray the strength of the association between the states of a child node conditional on the probabilities of the states of its parents (e.g. probability of drought risk (available soil water node) knowing that rainfall is above normal (rainfall amount node)).^{23,29} Specifying CPTs manually can be difficult particularly for complex nodes with many states and parents.²⁹ For example, a three states node with only three parents, each of which can assume 5 states, will require probability estimations for 375 possible situations. The `make_CPT` function from the `decision-Support` package in R provides a shortcut for deriving CPTs relating a child node to its parents.³⁰ The function requires the prior probability distribution of the child node, the child node sensitivity relative to the parents, the parents' effects, and the weight of influence of each parent to create the full CPT³⁰ using the likelihood method.²⁹ Estimates of all these parameters are elicited from the experts. The likelihood method is rooted in Bayes' theorem but focuses on the likelihood instead of conditional probability. The former is easier to estimate than the latter, because expert only

need to worry about a limited number of values to estimate.^{28,29} For a node of interest, the method assumes the conditional probability of a given child state to be the product of the prior probability of that child state and its likelihood given the states of the parents of the node. Since the prior can also be easily estimated by the expert, the method substantially facilitates expert elicitation. With further mathematical simplifications,²⁹ the expert only needs to provide few easier to estimate parameters for calculating the full CPT. These are the prior distribution of the child node, the base of the logarithm, a weight factor for the child node states, and a weight factor for the parent nodes states respectively corresponding to child prior, child node sensitivity, child state ranking, and the parents' effects in the make_CPT function.^{29,30}

III.3.2. The models and the modelling interface

The make_CPT routines were automated and interfaced with the cptable function from the gRain package,³¹ which provides architecture for computer readable graphical models, to formalise the BNs following the experts' causal reasoning. The posterior distributions of the 3645 parameters BNs was then used as synthetic inputs to feed the MC model using a routine that recursively accounts for each state of the node farming constraints at run time. This is equivalent to sampling over all possible farming constraint scenarios. Technically, we used MC particle filters^{32–34} to generate the probabilities of farming constraints conditional on different combination of the states of variables involved in the BNs. The farming constraint was then used as a loss factor along with the other quantitative yield metrics (see section III.1) to feed the MC model (Figure 2). The BNs were programmed to provide the 90% confidence probability distribution, the minimal, the maximal, and the median values of the farming constraint factor. Note that the probability distributions of the node farming constraints were chosen, at each stage of crop development, based on reasonable bounds for skewness and kurtosis via visual observation supported by bootstrapping.³⁵ We fitted several candidate distributions using the fitdist function from the fitdistrplus package³⁵ from which we chose the best fitting distribution. Note also that the remaining parameters (i.e. the minimal, maximal and median values) required by the decisionSupport's mcSimulation function,³⁰ which was used to conduct the MC simulation, were simply computed from the fitted distribution using the fitdist function and other facilities provided by the rriskDistributions package.³⁶ The same parameters were derived for the remaining MC nodes by the experts and cross-checked against their corresponding rainfed agriculture metrics for the same areas.

IV. Result

IV.1. Overview of the Conceptual Model

The overview of the model is presented in Figure 4 as mentioned in section III. For more a detailed description, the reader is referred to the supplementary material where we provided the model specifications at the initial stage of crop development (see Suppl.). The full model along with its complex specifications (121 and 32 nodes respectively for the BNs and MC models) are provided as reproducible code in the technical material. The usefulness and usability are presented as standalone case studies to showcase the model. The 121 nodes in the main BNs were grouped into 3 local BNs at each of the 4 crop development stages and used to describe 5 important FBFS processes. These processes were then used to describe the farming constraints which was considered as quantitative node to be part of the 32 nodes of the MC model. The further modelling process includes four work streams (Figure 4), represented by the individual variables (e.g. soil nutrients in pink), the local BNs (e.g. cropping options in green) describing the sub-processes formed by these variables,

the main BNs describing the farming constraints (in greenish yellow) to which these local BNs converge, and the crop development subject to the farming constraints (light green).

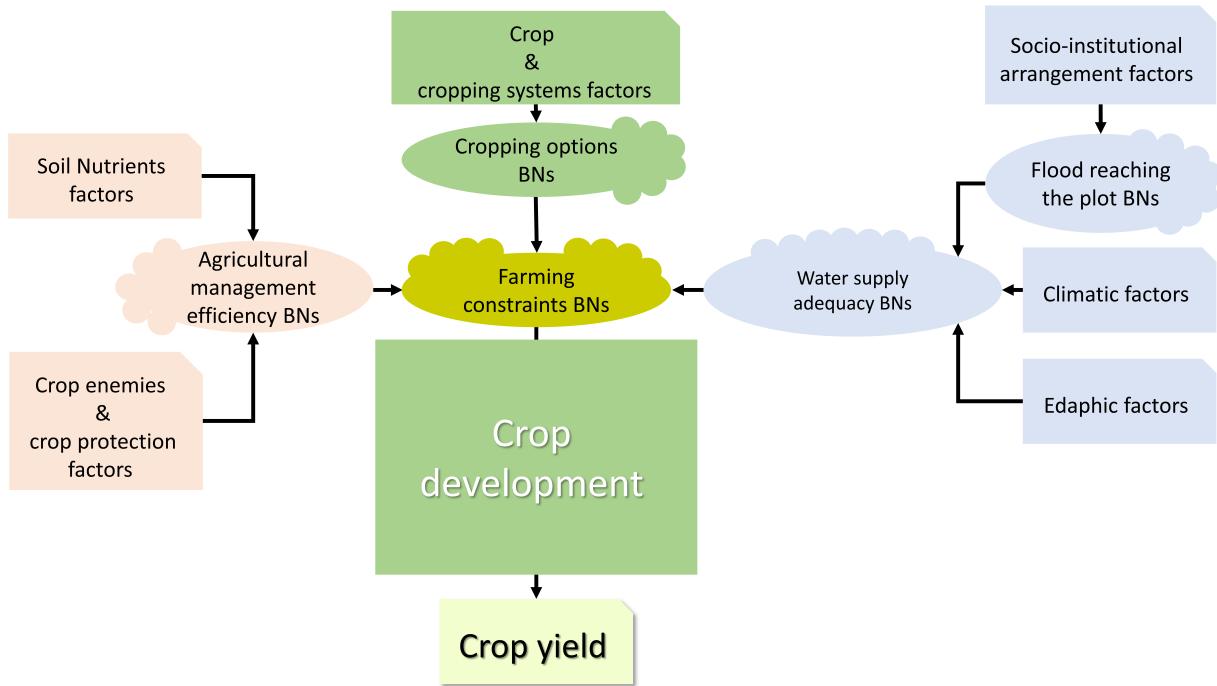


Figure 4: Conceptual model used to integrate the local BNs as part of a mixed model describing input allocation and agricultural management in FBFS in Kenya and Ethiopia.

IV.2. Case studies

IV.2.1. Qualitative Soil water Assessment under different Management Decisions at initial stage of crop development in Flood-based livelihood systems.

Abstract Add some text

Keywords: FBFS, Soil water content, Soil type, manure application, water conservation

Introduction FBFS are possible solutions to the increasing rainfall variability and drought in rainfed agriculture, because they can overcome water shortages through supplementary irrigation. To resolve potential drought risk, this irrigation must secure large amount of flood water in soil without creating a side effect of anaerobic conditions for crops. This requires maximizing water storage in soils (e.g. run-on, rainfall) while limiting waterlogging of soils and water loss (e.g. evapotranspiration, run-off, percolation). In FBFS, much of the water is acquired via flood, drought/waterlogging depend heavily on the amount of flood reaching the plot, and water losses are inversely related to the quality and quantity of soil organic amendments and other intrinsic factors related to soils along with other climatic factors. Depending on the soil type, farmers optimize the soil water at plot level by diverting the desired amount of flood, sometimes in combination with manure application and mechanic works. While most farmers use one or more of such soil and water conservation techniques, few can tell which the best option is. The objective of this case study is

to demonstrate how the model could be used to prescribe optimum pre-season cultural practices for improved soil water under specific soil types using the BNs sub-model. Towards this end, we queried the BNs sub-model targeting the node available soil water (AvailSoilWater) conditional on different combination of the states of 3 parent nodes that are relevant for soil water. These are the nodes amount of flood reaching the plot (FloodReachsPlot) which is a direct parent (Figure 5), the soil type (SoilType) and the manure application (ManureApp) which are ancestors (indirect parents).

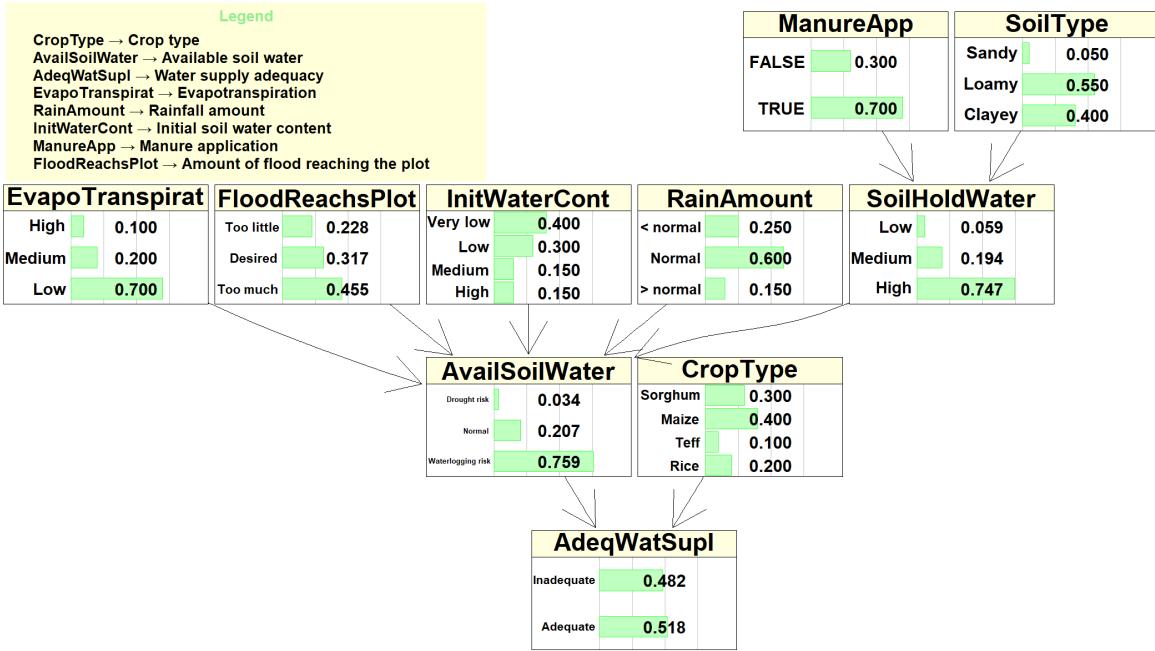


Figure 5: Illustration of the causality defining the factor of soil water content in FBFS in Kenya and Ethiopia.

Materials and Method A split-split plot experiment with 10 repetitions was used to test the effect of 3 different treatments and their interactions on the available soil water at initial stage of crop development (Figure 10; Suppl.) as node of interest. These include 3 different soil types (clayey, loamy, sandy) as the main plot, 2 manure applications levels (true, false) as subplot and 3 levels of amount of flood reaching the plot at initial stage as sub-subplot (too little, desired, too much). Split-split plot design extends the split plot design, both extensively described in 37; Chap. 3 - 4 and 38; Chap. 11. Each treatment corresponds to a combination of plot, subplot and sub-subplot. Such a design was considered as evidence based on which the BNs was queried targeting the node of interest. The resulting evidence-based posterior distributions were then fitted using an Analysis of Variance (ANOVA) Model followed by Tukey's honestly significant difference (Tukey HSD) post hoc test using the agricolae package in R. 39,40

Results and Discussions An ANOVA was conducted to test the effects of soil type, manure application and the amount of harvested flood water on the available soil water for crops at the initial stage of crop development using a split-split plot experimental design. It appears that

all treatments and their interactions are statistically significant ($p < 0.05$; Table 1; Suppl.) for the 3 states of the target node expect the interaction between the treatments Amount of flood reaching the plot and Manure application when the target is set to waterlogging ($p > 0.05$; Table 1; Suppl.). Blocks effect are not statistically significant ($p < 0.05$) owing to their negligible contribution in the experiment. The post hoc test (Figure 11; Suppl.) revealed that clay soils are better for soil water contrary to sandy ones. Even without manure addition and little amount of flood, drought is unlikely to occur under clayey soils. A similar situation is observed under sandy soils when it comes to waterlogging. While drought is not expected under both clayey and sandy soils supplemented with manure and good flood, the risk of waterlogging is eminent particularly under clayey soils. In general, normal conditions are rather rare as result of scoring probability below 0.5 except in the sandy soils which have low chance for waterlogging. In general, drought, contrary to waterlogging, is equally unlikely under both loamy and clayey regardless of the manure application if at least the desired amount of flood is expected .

IV.2.2. Probabilistic Assessment of Biomass Accumulation in Major Crops grown under FBFS in Kisumu, Kenya and Tigray, Ethiopia.

IntroductionApart from their substantial contribution to food security, FBFS also provide extensive amounts of fodder and post-harvest grazing opportunities for livestock in many areas. In both the Tigray region and Kisumu County, livestock feed on crop residues and other flood recession grasses, which make substantial contributions to livestock production. The symbiotic relationship between crops and livestock are well known and common to many farming systems. While the crop biomass contribution is substantial for livestock, biomass yield is often overlooked in favour of grain yield by many crop models. This case study provides a purely probabilistic attempt towards using the model for assessing the expected biomass yield and other important biomass yield metric (e.g. biomass yield gap, exploited biomass yield). By providing these estimates at different stages of crop development, the model may also serve as tool for monitoring biomass accumulation over the growing season.

Materials and MethodWe used the model as described in section III.3 to simulate the biomass yield at both the initial and late stage of crop development. In this lumped assessment, all possible scenarios across the different nodes specified in the BNs were considered. With the 3 states for each of the 4 nodes describing the farming constraints at different development stages (section III.1), and the 4 metrics considered (i.e. exploitable yield gap, exploitable yield gap loss due to constraints, actual exploited yield gap, and expected actual yield), the model simulation produced 324 ($324 = 3^3 * 4$) results corresponding to the space of possibilities. Only a few of these results are presented here to showcase the model. We selected the initial and the late stages of crop development along with the worst, the medium, and best cases scenario to respectively show the biomass accumulation over time, and the effect of varying farming constraints.

Results and DiscussionsThe simulated actual biomass yield along with other yield metrics are presented in Figures 12 and 13. The yellow and the gray colors respectively represent values inside and outside the 90% confidence interval. As it can be seen there are clearly important risks in FBFS practice in the study areas. In general yield / yield gain vary according to farming constraints with relatively higher chance of low biomass with increasing farming constraints. While the variability is substantial with time and varying level of farming constraints, the distributions have not changed much. Based on the 90% confidence intervals, the exploitable yield gap, the

actual exploited yield gap, and the expected yield can nearly double between high and low farming constraints contrary to the yield opportunity loss . These have also exhibited interesting change with time. Zero chances of yield opportunity loss, which seems to have tied distribution and little change with time, are plausible with minimal farming constraints. However, there seems to be more of possibilities for high gains than there are for high losses in terms of exploitable yield gaps, and this has not much changed with time. In a nutshell, based on the 90% confidence interval, there are possibilities for substantial increases in biomass yields biomass yield regardless of the level of the farming constraints. This increase, however, seems to be more important moving from high to medium level of constraints than it is between the medium and the low level.

IV.2.3. Quantitative grain yield variability and relative performance of Rice and Sorghum under different scenarios for water and crop enemies at mid-stage of crop development in flood-based Agriculture.

Introduction Many important risk factors for farmers; such as pest and diseases, weeds; are often overlooked by crop models because they are either difficult to measure or too complex to include . The impact of pest and disease or weeds on crop yield becomes even more important in FBFS due to their hydrology. In Rabour (Kisumu, Kenya) for example, rice grains can be destroyed by birds without what the famers refer to as the “bird scarring”. Depending on the spacing, some crops may require more of the farmer’s skill in the matter than others. For example, weed removal in teff can be difficult and time consuming compared to sorghum which is traditionally planted in relatively sparse rows. Weed seeds may be carried in by the flood water which further creates suitable conditions for their germination and growth. In most FBFS, therefore, farmers must protect their crops at every stage against a host of biotic stressors. This becomes challenging when the available labour force is limited, in which case the farmer has to prioritize certain crop protection activities over other. Such prioritization should be based on comparative analysis of plausible outcomes for different scenarios. The objective of this short case study is to demonstrate how the model can be used to provide a relative assessment of the impact of such aspects on crop yield. We considered a factorial experiment with 4 different treatments to study the effect of pests, diseases, weeds, and the available soil water on grain yield of rice and sorghum.

Materials and Methods To assess the effect of pests, diseases, weeds, the available soil water and crop type on grain yield, a 2x2x2x2 factorial experiment was considered. These 4 treatments were tested in a completely randomized block design (Figure 14. Suppl.) where repetitions are 1000 model runs of the mcSimulation function.³⁰ The rationale and the technical details on factorial/randomized block design can be found in [38](#) and [37](#). For each of the 4 treatments, we selected the 2 extremes states of the corresponding node in BNs to derive a sequence of all possible combinations. These were then used as evidence to query the posterior distribution of the node farming constraint in the BNs. Scenarios were assumed at the mid stage of crop development. The results of the BN queries were used to provide customized estimates of the node farming constraints to the MC model as described in section III.3.2. The simulated grain yield data were then submitted to ANOVA and Tukey HSD post hoc testing using the agricolae package in R.^{[39](#),[40](#)}

Results

Main effects of the different treatments on grain yield

The between-subjects effect from the ANOVA test are reported in Table 2; Suppl. In these results, the main factors of crop yield are pest and diseases ($p < 0.01$; Table 2; Suppl.), weeds ($p < 0.05$; Table 2; Suppl.), and the crop type ($p < 0.01$; Table 2; Suppl.). It is unclear how the available soil water affects crop yield ($p > 0.05$; Table 2; Suppl.). However, statistically significant differences are recorded when other factors, such as weeds ($p < 0.05$; Table 2; Suppl.) or pest and diseases, are considered ($p < 0.05$; Table 2; Suppl.). Pest and diseases accentuated the impact of weeds on crop yield and crop yield variability more than did the water regime ($p < 0.05$; Table 2; Suppl. 5 and 9). The reverse is observed when the crop type is considered (Table 2; Suppl. 11, 14). This means pest and diseases are strongly related to other factors, hence their implications on grain yield are to be explained in the context of other factors for which mutual influences can be expected.

Marginal and Joint effects of the levels of the different treatments on grain yield

The Tukey's honestly significant difference (Tukey HSD) suggests significant differences in the levels of the different treatments ($p < 0.01$; Table 3; Suppl.) except in the case of available soil water which was not statistically significant in the ANOVA test. Crop type seems to be the most important factor of grain yield variability followed by the available soil water. Weeds impact seems to be the least important. In general, Rice outperformed Sorghum (diff = 6.76; $p < 0.01$; Table 3; Suppl. 33). When the farmer planned for both crops, waterlogging is worse than drought risk, particularly under well managed weeds (diff = -15.28 $p < 0.01$; Table 3; Suppl. 25). When proper pest & diseases control is considered, waterlogging is better than drought conditions coupled with severe level of pests & diseases (diff = -11.06; $p < 0.01$; Table 3; Suppl. 34). Rice under drought is better than Sorghum under waterlogging condition (diff = -92.02; $p < 0.01$; Table 3; Suppl. 25) regardless of the state of crop protection deployed for sorghum (diff \geq -55.53; $p < 0.01$; Table 3; Suppl. 4 2 - 3). However, Sorghum under waterlogging outperformed Rice subjected to drought conditions (diff \geq -55.53; $p < 0.01$; Table 3; Suppl. 4 4 - 5) suggesting the importance of water management strategy employed by the farmer. Under the same scenario, increasing the effect of pest & disease along with the effect of weeds in rice resulted in important decrease (up to 3 times) of the relative performance of Rice (diff = -92.02; Table 3; Suppl. 1 versus diff = -34.03; Table 3; Suppl. 8; $p < 0.01$) testifying the importance of the regularity and skills of the farmers in crop protection on crop yield. The overall comparative analysis from of the joint effect of the treatments from the post hoc test suggest rice coupled with proper crop protection methods to be the best option when drought risk is expected towards the mid stage of crop development (Table 3; Suppl. 1 & 4). When waterlogging is expected, however, no big difference can be expected between the two crops (Table 3; Suppl. 4 2-3). In the case of waterlogging, neither pest & disease control nor the weeds removal seems to promote sorghum over rice (Table 3; Suppl. 4 2-5). When Sorghum is grown under waterlogging without proper weed control, no big difference is observed with better pest & diseases control (Table 3; Suppl. 4 5-6) suggesting that either Sorghum is insensitive to pest & diseases, or pest & disease in Sorghum are insensitive to the protection methods employed by the farmers, or the effect of waterlogging and /or weeds mask the effect of pest & disease. This unclear behaviour of pest and diseases is also observed under rice subjected to drought (Table 3; Suppl. 4 7-9). Under drought conditions, Rice is no better than Sorghum even when proper weeding is considered (Table 3; Suppl. 4 10-11). Weeds control seems equivalent to pest & diseases control in drought-affected sorghum (Table 3; Suppl. 4 13-14) which performed equally under waterlogging coupled with suitable crop protection (Table 3; Suppl. 4 7-15).

V. Discussion

For example, crop yield in certain crops such as maize are very sensitive to water stress during flowering and cob formation than during other development stages (Pault, 2013).

reference

1. Murthy, V. R. K. Crop Growth Modeling and Its Applications in Agricultural Meteorology. *Satellite Remote Sensing and GIS Applications in Agricultural Meteorology* 235–261 (2004).
2. Boote, K. J., Jones, J. W. & Pickering, N. B. Potential uses and limitations of crop models. *Agronomy Journal* **88**, 704–716 (1996).
3. Cerdan, O., Souchère, V., Lecomte, V., Couturier, A. & Le Bissonnais, Y. Incorporating soil surface crusting processes in an expert-based runoff model: Sealing and transfer by runoff and erosion related to agricultural management. *Catena* **46**, 189–205 (2002).
4. Luedeling, E., Whitney, C., Rosenstock, T. & Shepherd, K. "Future Agriculture : Socio-ecological transitions and bio-cultural shifts". in *Modelling agricultural realities to support development decisions* (ed. Tielkes, E.) 248 (Tropentag Bonn, 2017).
5. Luedeling, E. & Shepherd, K. Decision analysis principles can guide the modelling of complex agroforestry systems. in *International crop modelling symposium* (eds. Ewert, F., Boote, K. J., Rötter, R. P., Thorburn, P. & Nendel, C.) 308–309 (iCROPM2016, 2016).
6. Rötter, R. & Höhn, J. An overview of climate change impact on crop production and its variability in Europe, related uncertainties and research challenges. in *Climate change and food systems: Global assessments and implications for food security and trade* (ed. Elbehri, A.) 43 (FAO, 2015).
7. Van Ittersum, M. K. *et al.* Yield gap analysis with local to global relevance-A review. *Field Crops Research* **143**, 4–17 (2013).
8. Haile, A. M. *A Tradition in Transition, Water Management Reforms and Indigenous Spate Irrigation Systems in Eritrea*. 212 (CRC Press, 2010).
9. Steenbergen, F. van, Lawrence, P., Mehari, A., Salman, M. & Faurès, J.-M. *Guidelines on spate irrigation*. 249 (FAO, 2010).
10. Junk, W., Bayley, P. & Sparks, R. The Flood Pulse Concept in River - Floodplain Systems. in *Proceedings of the international large river symposium* (ed. Dodge, D. P.) 110–127 (Can. Spec. Publ, 1989).
11. Harlan, J. R. & Pasquereau, J. Décrue Agriculture in Mali. *Economic Botany* **23**, 70–74 (1969).
12. FBLN. Flood-based Livelihood Network (FBLN) Foundation. (2018).
13. Puertas, D. G.-L., Steenbergen, F. van, Haile, A. M., Kool, M. & Embaye, T.-a. G. Flood based farming systems in Africa. 52 (2011).
14. Steenbergen, F. van, MacAnderson, I. & Mehari, A. Spate irrigation in the Horn of Africa: Status and potential. 44 (2011).
15. Erkossa, T., Langan, S. J. & Hagos, F. Constraints to the development, operation and maintenance of spate irrigation schemes in Ethiopia. in *Proceedings of the workshop on flood-based farming*

for food security and adaption to climate change in ethiopia: Potential and challenges (ed. Erkossa, Teklu; Hagos, Fitsum; Lefore, N.) 5–22 (International Water Management Institute, 2014).

16. Hardaker, J. B., Anderson, J. R., Lien, G. & Huirne, R. B. *Coping with Risk in Agriculture: Applied Decision Analysis*. 289 (CABI Publishing, 2015).
17. Jax, K. & Setälä, H. Function and "Functioning" in Ecology: What Does It Mean? *Oikos* **111**, 641–648 (2005).
18. Allen, R. G., Pereira, L. S., Raes, D. & Smith, M. Crop evapotranspiration: Guidelines for computing crop requirements. 300 (1998). doi:[10.1016/j.eja.2010.12.001](https://doi.org/10.1016/j.eja.2010.12.001)
19. Pearl, J. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. 573 (Morgan Kaufmann Publishers, Inc., 1988).
20. Fenton, N. & Neil, M. *Risk Assessment and Decision Analysis With Bayesian Networks*. 493 (CRC Press, 2013).
21. Jensen, F. V. *An Introduction of Bayesian Network*. 178 (Springer; Har/Dskt edition, 1996).
22. Hubbard, D. W. *How to Measure Anything: Finding the Value of Intangibles in Business*. 432 (John Wiley & Sons, Inc., 2014).
23. Scutari, M. & Denis, J.-B. *Networks Bayesian With Examples in R*. 241 (CRC Press, 2015).
24. Rosenstock, T. S. *et al.* Targeting conservation agriculture in the context of livelihoods and landscapes. *Agriculture, Ecosystems and Environment* **187**, 47–51 (2014).
25. Jensen, F. V. & Nielsen, T. D. *Bayesian Networks and Decision Graphs*. 45, 448 (Springer, 2007).
26. Luedeling, E. *et al.* Fresh groundwater for Wajir - ex-ante assessment of uncertain benefits for multiple stakeholders in a water supply project in Northern Kenya. *Frontiers in Environmental Science* **3**, (2015).
27. Whitney, C. *et al.* Probabilistic Decision Tools for Determining Impacts of Agricultural Development Policy on Household Nutrition. *Earth's Future* **6**, 359–372 (2018).
28. Whitney, C., Shepherd, K. & Luedeling, E. Decision analysis methods guide; Agricultural policy for nutrition. 27 (2018). doi:[10.5716/WP18001.PDF](https://doi.org/10.5716/WP18001.PDF)
29. Hansson, F. & Sjökvist, S. Modelling Expert Judgement into a Bayesian Belief Network -A Method for Consistent and Robust Determination of Conditional Probability Tables. 53 (LUND University, 2013).
30. Luedeling, E. & Goehring, L. decisionSupport: Quantitative Support of 1.103.7., Making under Uncertainty. R package version 1.103.7. (2018).
31. Højsgaard, S. Graphical Independence Networks with the gRain Package for R. *Journal of Statistical Software* **46**, 1–26 (2012).
32. Kitagawa, G. Monte Carlo Filter and Smoother for Non-Gaussian Nonlinear State Space Models. *Journal of Computational and Graphical Statistics* **5**, 1–25 (2016).
33. Koller, D. & Friedman, N. *Probabilistic Graphical Models: Principles and Techniques*. 1270 (MIT Press, 2009).
34. Scutari, M. Learning Bayesian Networks with the bnlearn R Package. *Journal of Statistical Software* **35**, 1–22 (2010).

35. Delignette-Muller, M. L. & Dutang, C. fitdistrplus: An R Package for Fitting Distributions. *Journal of Statistical Software* **64**, 1–34 (2015).
36. Belgorodski, N., Greiner, M., Tolksdorf, K. & Schueller, K. rriskDistributions: Fitting Distributions to Given Data or Known Quantiles. R package version 2.1.2. (2017).
37. Gomez, K. A. & Gomez, A. A. *Statistical procedures agricultural research.* **6**, 627 (John Wiley & Sons, Inc., 1984).
38. Panse, V. G. & Sukhatme, P. V. *Statistical Methods for Agricultural Workers.* 361 (Indian Council of Agricultural Research, 1957).
39. De Mendiburu, F. Package ‘agricolae’. (2016). doi:[10.2307/2411227](https://doi.org/10.2307/2411227).
40. R Core Team. *R: A language and environment for statistical computing.* (R Foundation for Statistical Computing, 2018).