

Modelling Risk and Uncertainty in Flood-based Farming Systems using a Knowledge-based Approach.

Issoufou, Liman Harou^{a,b}, Cory Whitney^c, James Kungu^a, Eike Luedeling^c

^a*Kenyatta University, Department of Environmental Sciences, P.O. Box 43844 00100 Nairobi, Kenya*

^b*World Agroforestry Centre (ICRAF), United Nations Avenue, Gigiri, P.O. Box 30677-00100, Nairobi, Kenya*

^c*University of Bonn, Department of Horticultural Sciences, Auf dem Hügel 6, D-53121, Bonn, Germany*

^d*Center for Development research (ZEF), University of Bonn, Genscherallee 3, D-53113, Bonn, Germany*

Abstract

Crop models support agricultural decisions under various agricultural conditions. However, problems related to the scope of application, data availability, system complexity, or system-specific settings make many models applicable only with far-reaching and often questionable assumptions. In this paper, we used Decision Theory and participatory methods to address these challenges in a complex system where data and resources are limited. The resulting generic models is customizable and solution-oriented, and it accounts for all important variables and their interactions, as determined by local and foreign experts as well as relevant literature. We describe how such a probabilistic model can be developed and customized to specific situations based on case studies related to flood-based farming systems in Ethiopia and Kenya. Bayesian networks, describing stochastic system components, are used to generate inputs for a Monte Carlo model that simulates plausible ranges of expected grain and biomass yields at various stages of crop development.

Introduction

Crop modelling has been used to provide guidance to scientists and decision makers on various aspects of agriculture. Several of these crop models were developed to assess aspects of high agricultural importance (e.g. soil water, soil carbon, crop growth, crop yield under constraints). Despite their usefulness and widespread use across the globe, there are often concerns about several aspects such as the requirement for precise data which are largely unavailable or the complexity of the system being model. This work attempts to address these concerns by using a set of Decision Theory methods to propose a new approach to crop modelling. We use all available sources of information to develop a generic crop model for flood-based farming systems (FBFS) by fusing Bayesian networks (BNs) and Monte Carlo models. BNs are used to include qualitative variables, whereas MC models were used for quantitative ones. We describe how such a model can be developed and provide ways for deriving customized models for specific situations using three case studies. Each case study presents a customized instance of the generic model to demonstrate the usefulness and usability of the model from different perspectives. The modelling exercise took place in Kisumu County in

Email addresses: issoufoul@gmail.com; L.issoufou@cgiar.org (Issoufou, Liman Harou), cory.whitney@uni-bonn.de (Cory Whitney), kungu.james@ku.ac.ke (James Kungu), luedeling@uni-bonn.de (Eike Luedeling)

Kenya and the Tigray region in Ethiopia. This document provides supplementary materials supporting the main paper. The details provided here concern, mainly, the different modules and sub-module that make up the generic model. Still, this material will only give a general idea about the model. The full model is provided as reproducible code in two repositories available on GitHub. The first one is a package describing the different functions. The second is an Rmarkdown folder containing the necessary ingredient to reproduce the main paper and this supplementary material.

1. Description of the modules in the BN models

1.1. Farming constraints at farming plot level

The purpose of the BNs is to describe the node ‘farming constraints’ (FarmConstraints), which accounts for all limiting factors deemed important to FBFS in the study areas. The most important limiting factors describing the farming constraints in the study areas (Figure 1) are the adequacy of water supply to crops (AdeqWatSupl), the effectiveness of cropping systems (EffCropSystems), and the quality of farming practices (AgricManagEff) resulting from farmers’ managements options (Figure 2 to Figure 6). These were described in three modules: the soil water, the cropping system and the management modules. They are described in more detail in sections 1.2 to 1.4.

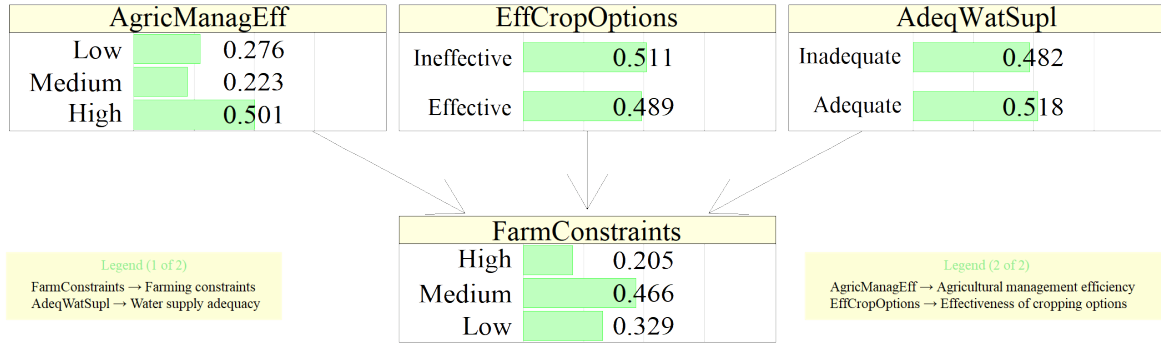


Figure 1: Bayesian Network describing important variables as part of a mixed model describing the farming constraints in FBFS in Kenya and Ethiopia.

1.2. Water supply adequacy

Because of the complexity of water supply in FBFS, water supply adequacy was described via 2 sub-modules (Figure 2; Figure 3). The amount of flood reaching the farming plot (FloodReachsPlot) was assessed separately, since it is an important variable with regard to the overall available water for crop development in FBFS settings.

At the root of the BN that describes the amount of flood reaching the plot are the nodes describing water diversion type (DiversionType), the social arrangement for water acquisition and sharing (SocialArrang),

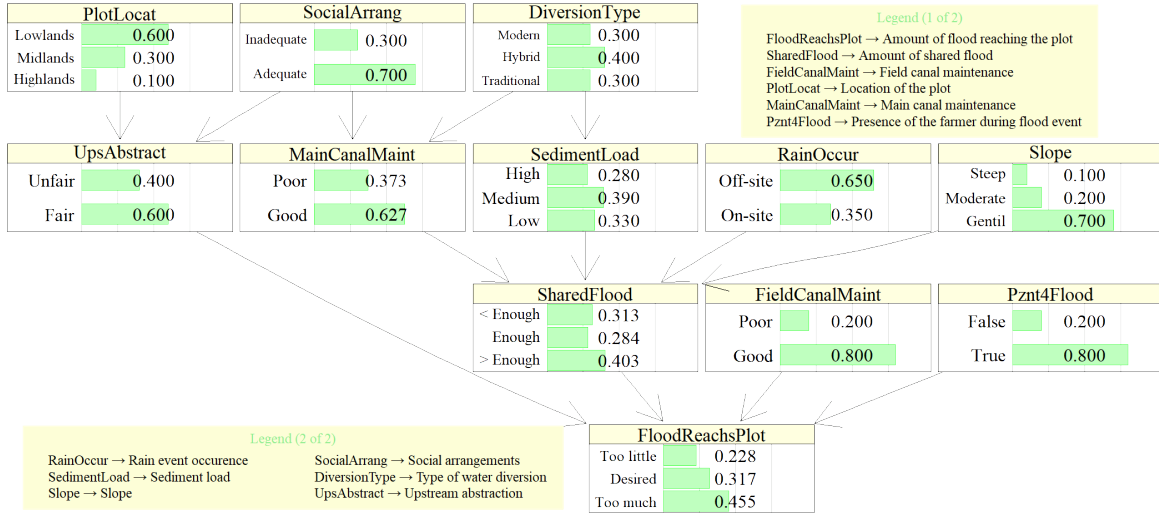


Figure 2: Bayesian Network describing important variables as part of a mixed model describing the amount of flood at the farming plot level in FBFS in Kenya and Ethiopia.

and the location of the farming plot (PlotLocat). The water diversion type is important, since it determines the overall water delivery (SharedFlood) and ultimately the overall expected flood amount at the plot level (FloodReachsPlot). It acts via the way in which farmers are committed as a group (SocialArrang) to maintain their main water distribution networks (MainCanalMaint) and its efficiency with regard to sediment load (SedimentLoad). If the social organisation (SocialArrang) is not adequate along a topographic gradient (Slope), this may lead to unfair upstream abstraction (UpsAbstract), hence penalising water users downstream (FloodReachsPlot). Flood water is trapped in FBFS fields using earthen bunds, which need to be opened at the right time either for irrigation or for water release if there is too much water. Individual farmers, therefore, need to make sure this happens by being present at the right time and making sure the water is diverted from the main canal to the respective secondary and tertiary sub-divisions of their water distribution system. Given the adequacy of the above-mentioned aspects, the presence of the farmers in the field during floods (Pznt4Flood) and the quality of canal maintenance (FieldCanalMaint) can still strongly affect the amount of flood reaching the plot (FloodReachsPlot), and ultimately the overall soil water budget. The presence of the farmer and the quality of canal maintenance can either have minor importance when it rains on the field (RainOccur = On-site) or major importance in the case of spontaneous flood events from upstream (RainOccur = Off-site).

The soil water profile of a given farming plot includes important biophysical aspects related to soils and climate. The available water for crops is determined by the soil water holding capacity (SoilHoldWater), the total amount and temporal distribution of rainfall (RainAmount) and residual soil moisture (InitWaterCont), which can be significant under certain FBFS such as flood recession agriculture. Soil water holding capacity (SoilHoldWater) can vary greatly depending on the soil type (SoilType) and the soil management strategies

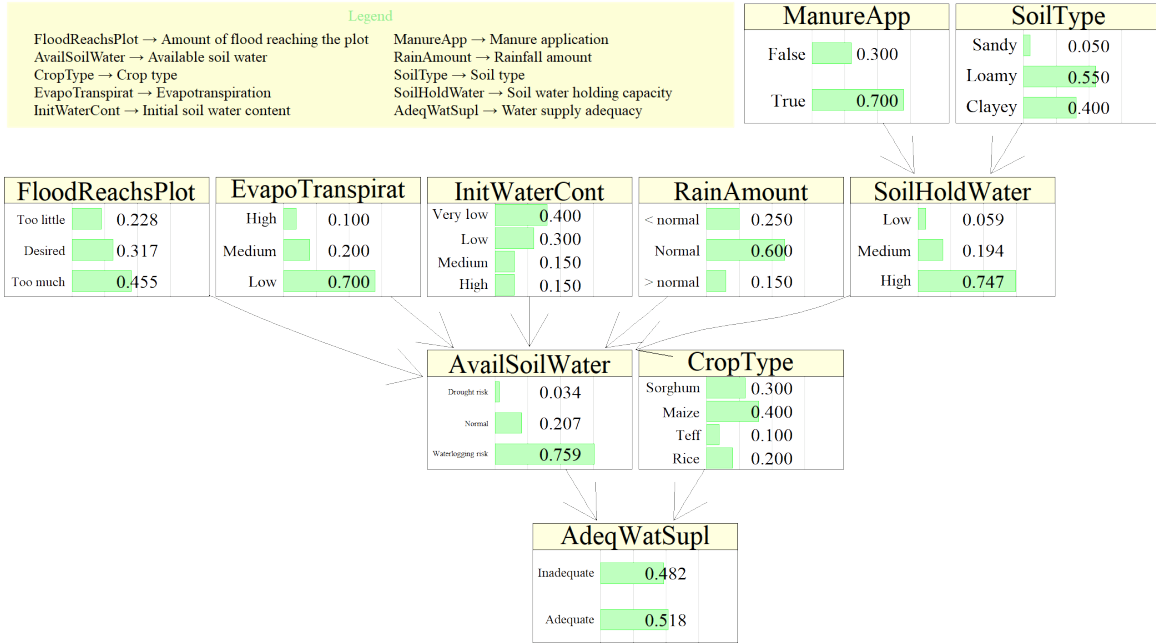


Figure 3: Bayesian Network describing important variables as part of a mixed model describing the available soil water content in FBFS in Kenya and Ethiopia.

employed by the farmer (e.g. manure application; ManureApp).

1.3. Agricultural management efficiency

Agricultural management efficiency (AgricManagEff) is described using two submodules (Figure 4; Figure 5): one describing the availability of soil nutrients (AvailSoilNut), and another describing the overall agricultural management efficiency. The availability of soil nutrients is defined by fertilizer applications either in the form of organic (ManureApp) or mineral additions (FertiApp) by the farmer, and on the other hand by the sediment naturally brought by the flood water. Here, it is important to distinguish between coarse sediments (SedimentLoad) described under the rubric Available soil water, which are likely to have negative impacts, and the fine rich sediments (AddSediments), which have positive impacts on crops. Fertilizer application can only be expected when the fertilizer desired by the farmer is available on the market at the time when the farmers can afford it (Access2Inputs). This depends on the purchasing power of farmers and their willingness to acquire the fertilizers (RelativeWealth), as well as on mutual assistance (MutualAids) among farmers.

The factors of agricultural management efficiency include the available soil nutrients (AvailSoilNut), crop exposure/sensitivity to pests and diseases (PestDiseaseImpact) and the severity of weeds (WeedsImpact). The magnitude and recurrence of damage to crops due to pests & diseases and yield losses due to weeds can be related to the skills of the farmers (FarmerSkills) in crop protection, the farmers' ability to acquire chemicals, and other aspects such as the available labour force, which may influence the effectiveness of crop protection.

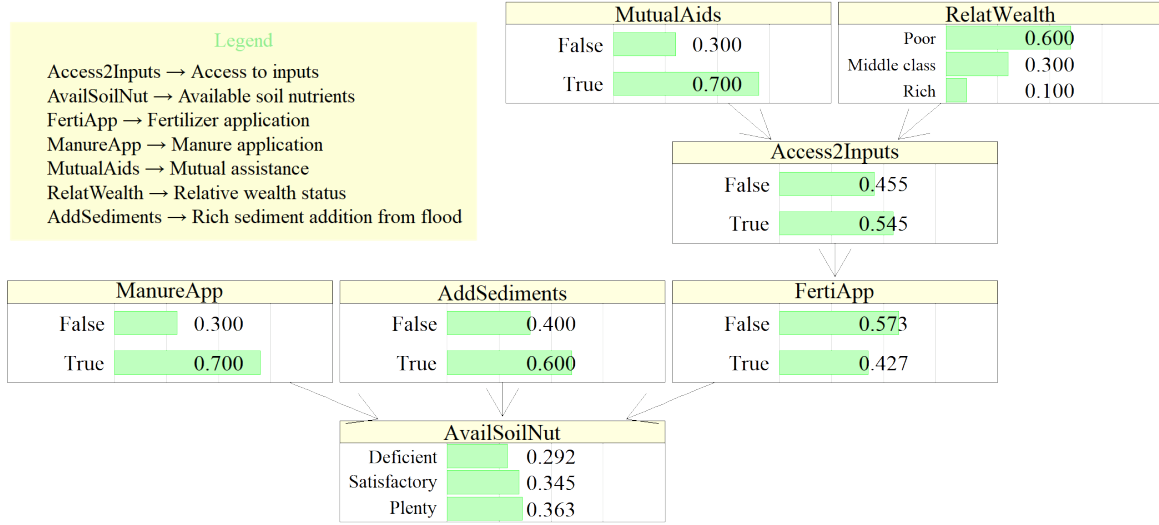


Figure 4: Bayesian Network describing important variables as part of a mixed model describing the available nutrients in FBFS in Kenya and Ethiopia.

1.4. Effectiveness of cropping options

The ‘effectiveness of cropping options’ (EffCropsOptions) BN describes the various factors shaping the cropping system at the very early stages of crop growth. Crops are defined by their type (CropType: rice, maize, sorghum, teff), their variety (CropVariety) and the farmers’ decision regarding which crops to grow and how to grow them (Intercropping). The decision of the farmer to grow a given type of crop and variety have strong yield implications.

2. BN structures and node sensitivity in relation to crop development

The local BNs described above were connected to form a single BN accounting for the complexity of the farming system. This BN aims to integrate the joint effect of all farming constraints in one single variable (see section 1.1) to be used along with other variables described in the MC model. In the MC model, the entire BN was replicated with some changes in the CPTs and the causal relations when necessary. For example, the impact of weeds may depend on the crop development stage. The model assumes that crop sensitivity to biotic and abiotic factors depends on the period when these factors are considered. Therefore, these stage-specific BNs or rather the concepts embedded in their nodes, their causal relationships, and their CPTs were developed according to available information. BNs were developed to estimate the farming constraint at plot level and at each phenological stage of crop development. The BN nodes (Figure 1 to Figure 6) are represented in the model in four crop development stages (e.g. amount of shared flood at initial stage, development stage, mid stage, and late stage). A few BN nodes that are expected to remain constant (e.g. the type of water diversion (DiversionType), the slope etc.), however, were represented only once in the model and transported through the various crop development stages. For example, the BN nodes ‘soil type’ and

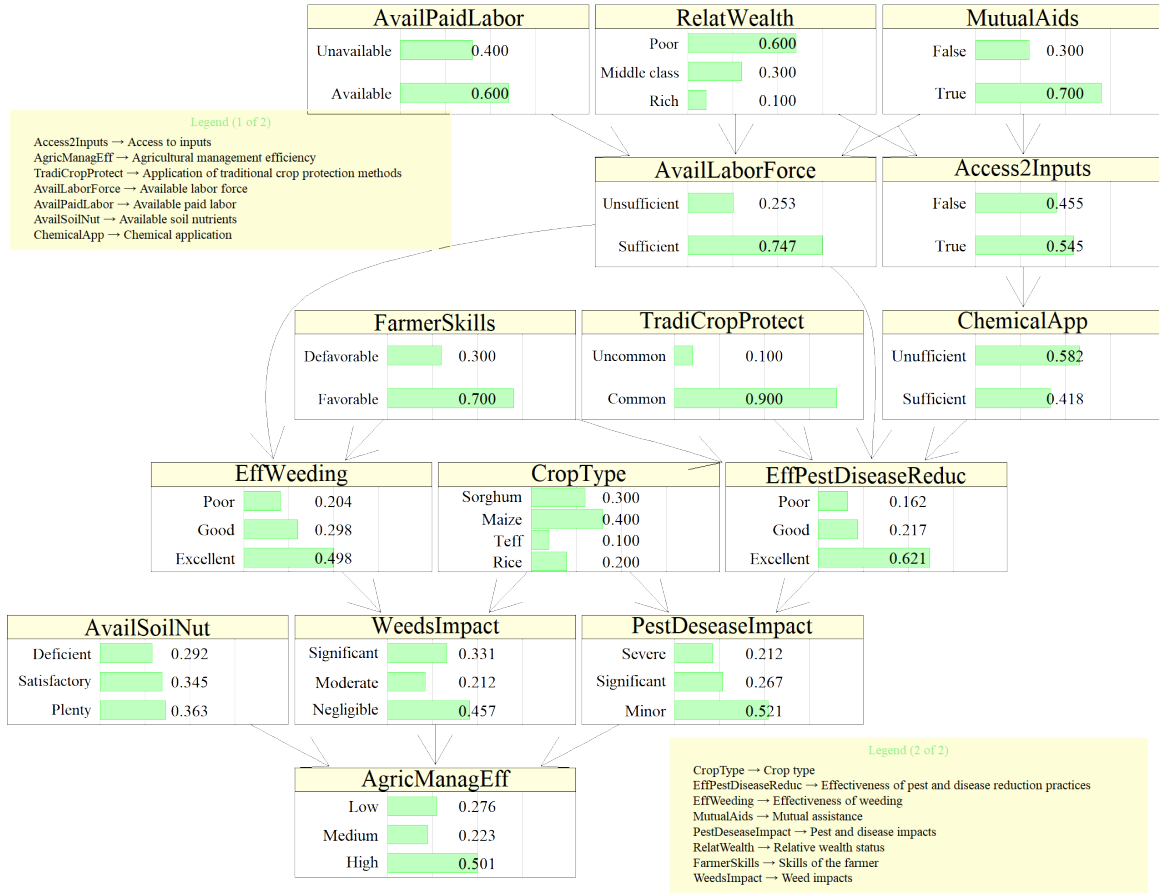


Figure 5: Bayesian Network describing important variables as part of a mixed model describing the effectiveness of agricultural management in FBFS in Kenya and Ethiopia.

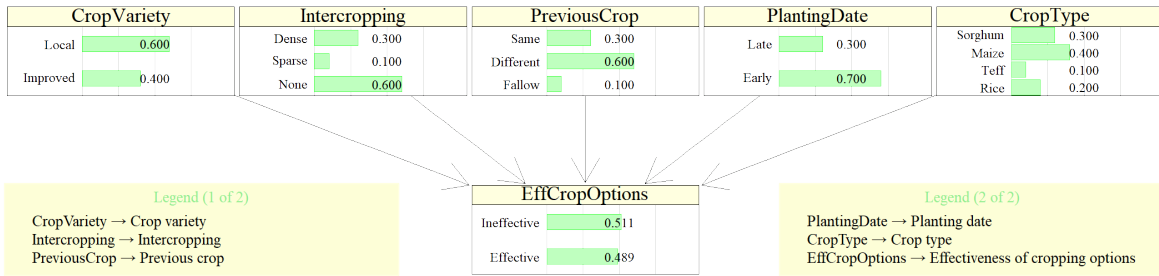


Figure 6: Bayesian Network describing important variables as part of a mixed model of the effectiveness of cropping options in FBFS in Kenya and Ethiopia.

‘location of the farming plot’ are considered static variables over time. They are unlikely to change, at least not within a few years, in contrast to other nodes such as the impact of weeds and diseases, which differs between consecutive phenological stages. We focus on processes at the initial stage in the description, but the

target variable is assessed quantitatively at each of the four stages to feed the MC model.

3. Quantitative assessment of Farming Constraints

The inclusion of the farming constraint variable in the MC models requires its quantitative description capturing all possible scenarios for the variable. For each scenario, these estimates include the range of variable values spanning the 90% confidence interval and its probability distribution as required by the mcSimulation function (Luedeling et al., 2019, see also the main paper). We generated the probabilities of the variable from the BN based on the query of interest. The resulting data were used to compute the lower bound, the median and the upper bound of the 90% confidence interval, and to estimate the shape of the best-fitting probability distribution given several candidate distributions and the ranges of skewness and kurtosis (Cullen and Fray graph, also skewness-kurtosis plot) of the generated data (Cullen and Frey, 1999; Delignette-Muller and Dutang, 2015). The probability distribution of the variable at the four crop development stages was assessed via visual observation of the Cullen and Fray graph (see example for the initial stage of crop development in Figure 7).

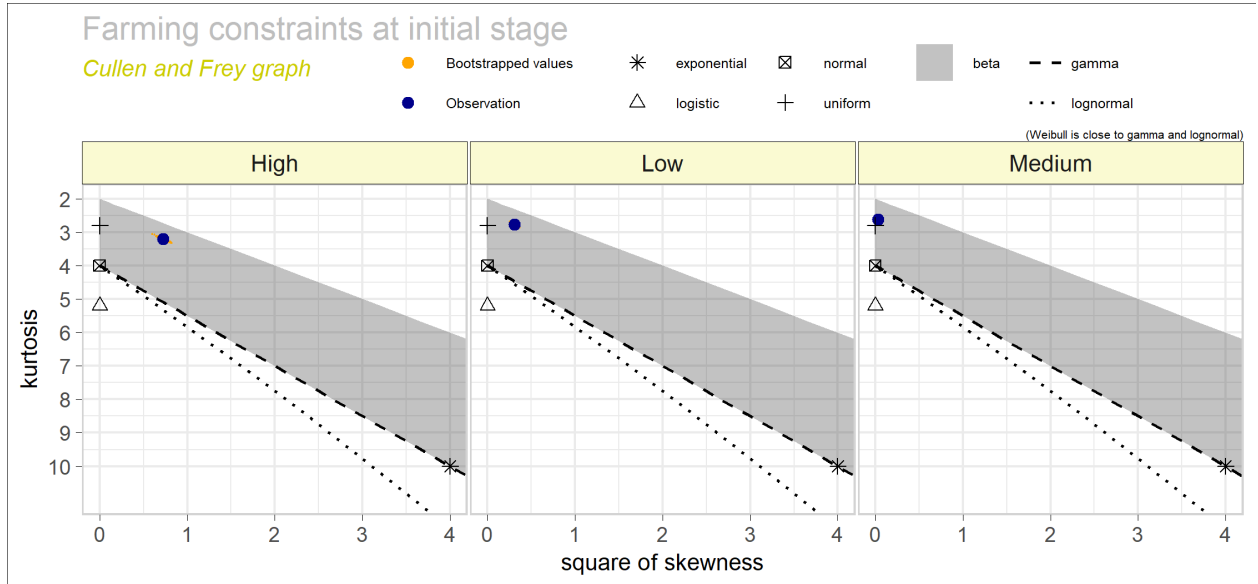


Figure 7: Cullen and Frey graph showing the bounds for skewness and kurtosis and the best fitting distribution of three states of the variable ‘farming constraint’ derived from the joint distribution of a Bayesian network used to feed a Monte Carlo model as part of a mixed crop model for flood-based farming systems in Kisumu (Kenya) and Tigray (Ethiopia). The Bayesian network describes the variable farming constraint at 4 crop development stages (i.e. initial, development, mid-season, and late stages). This figure provides an example for the initial stage.

The best fit for candidate distributions of the target node (farming constraints) derived from the sampling of the posterior probabilities of the BNs are shown in Figure 7. In these examples (corresponding to the case study on the probabilistic assessment of crop biomass in section 5.2), the distribution of the farming constraint variable at the initial stage of crop development was fitted to beta, uniform, and uniform distributions, respectively, for the three states of the variable (i.e. high, medium, low) owing to the values of the bootstrapped

Farming constraints at initial stage

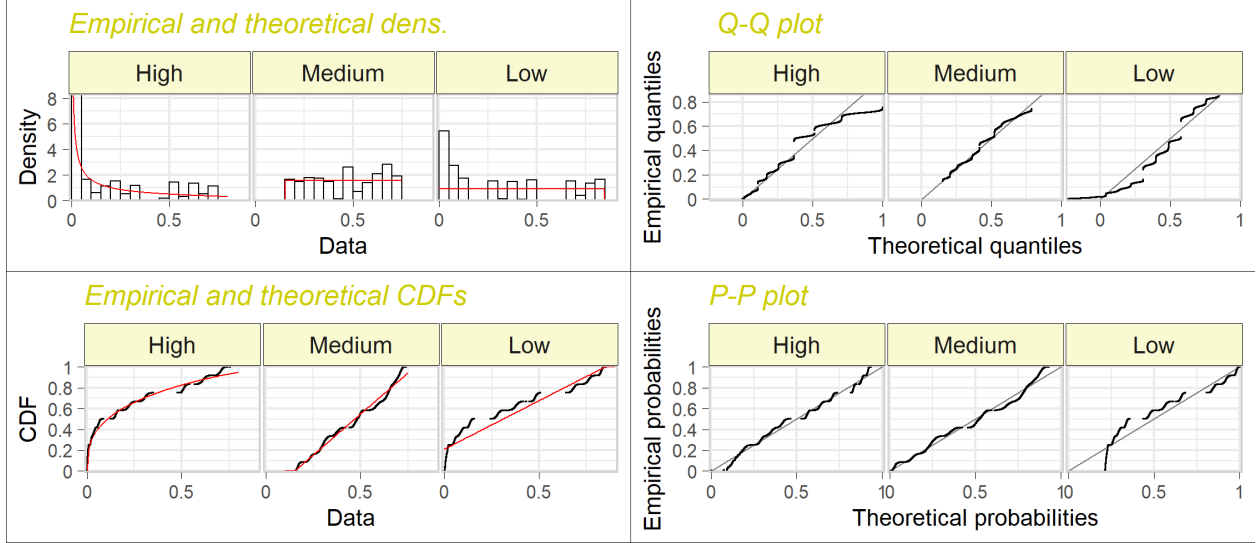


Figure 8: Graphical assessment of probability distribution fits, including the quantile and probability plots along with the theoretical and empirical density and cumulative distribution functions, for the variable farming constraint derived from the joint distribution of a Bayesian network used to feed a Monte Carlo model as part of a mixed crop model for flood-based farming systems in Kisumu (Kenya) and Tigray (Ethiopia). The Bayesian network describes the variable farming constraint at 4 crop development stages (i.e. initial, development, mid-season, and late stages). This figure provides example for the initial stage.

samples lying within reasonable bounds for skewness and kurtosis characterising these distributions. A graphical assessment of the fit including the quantile and probability plots, along with the theoretical and empirical density and cumulative distribution functions, is presented in Figure 8.

4. Model Carlo model

Several MC nodes were considered to quantitatively describe crop yield following the crop development subdivided into 4 stages. Figure 9 presents the graphical model including the initial and development stages. Note that the mid and the late stages were described using the same concept as in the case of development stage. Potential yield, average actual yield and biomass expansion factor were modelled using gamma distributions, whereas the exploitable yield potential and the harvest index were modelled using uniform distributions.

We modelled biomass accumulation following the crop development, and ultimately converted biomass into grain yield using a realistic range of possible harvest index values. The maximum attainable biomass yield (Attainable yield potential) depends on the crop potential yield (Yield potential) and its achievable limit (Exploitable yield potential). This concept acknowledges that it is impossible for farmers to fully eliminate the farming constraints. The exploitable yield gap is then estimated by subtracting the long-term average yield (Average actual yield) from the attainable yield. Yield opportunity losses (Exploitable yield gap), therefore, are proportional to the difference in input allocation among farmers (Farming constraints) relative to the exploitable yield gap. Therefore, additional yield opportunity for a given farmer (Actual exploited yield gap)

end, this instance of the model consists in querying the soil water module (see **Figure 5** in the main paper material) targeting the node describing the available soil water, conditional on various combinations of the states of 3 parent nodes that are relevant for soil water. We formulated all possible scenarios, accounting for each set of child-parent state relationships to ultimately investigate the probability of obtaining a specific soil water level. The emerging insights can be used diagnostically to prescribe optimal pre-season management for FBFS. The target node is a child node of nodes describing the amount of floodwater reaching the plot, the level of evapotranspiration, the initial water content and the amount of rainfall. The available soil water node is also dependent on the soil’s water holding capacity, which in turn is a child of manure application and soil type. These were considered only at the initial stage of crop development as a basis for querying the soil water module. The resulting posterior distributions were then used to study the probability of obtaining each of 3 levels of available soil water (drought risk, normal, waterlogging risk).

5.2. Probabilistic assessment of crop biomass in flood-based farming systems

Apart 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 in 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 is well known and common to many farming systems. While crop biomass makes a substantial contribution to livestock production, most crop models focus their attention on grain yields, often failing to consider biomass yield at all. This case study provides a purely probabilistic attempt towards using the crop growth module for assessing the expected biomass yield and other important biomass yield metrics (e.g. biomass yield gap, exploited biomass yield). By providing these estimates at different stages of crop development, this instance of the generic model may also serve as a monitoring tool for biomass accumulation over the growing season.

The biomass expansion model was specified as described in the main paper and used to simulate the biomass yield along with several important yield metrics (see sections on **Models for quantitative variables**, and **Crop growth module** in the main paper material) at both the initial and late stages of crop development. The **Figure 7** in the main paper summarises the MC model including the relationships between the quantitative variables and the mathematical calculations involved at the initial and development stages. The Bayesian networks describing the qualitative variables along with the full MC model are presented in the supplementary and technical materials. In this lumped assessment, all possible scenarios across the different nodes specified in the BN were considered. With the 3 states for each of the 4 nodes describing the farming constraints at the 4 crop development stages (section 5.2 and section on **Models for quantitative variables** in the main paper material), 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 ($324 = 3 \times 3 \times 3 \times 4$) results corresponding to the total set 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, medium, and best-case scenarios to show the biomass accumulation over time, and to illustrate the effect of varying farming constraints on crop development.

5.3. *Impact of major constraints on grain yield variability of rice and sorghum*

Many important risk factors that affect farms (e.g. pests, diseases and weeds) are ignored by many crop models, because they are either difficult to measure or too complex to include. The stochastic nature of many risk factors makes them difficult to predict with accuracy. In many agricultural settings, water, weeds, and pests & diseases rank among the top concerns and are perceived by farmers as major impediments to agricultural production. This is particularly important for FBFS, where the type of water management can strongly affect yield losses due to pests, diseases and weeds. Crop spacing can have a strong influence on the skills required of farmers for adequate management of biotic stressors. 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 can be washed onto fields by the flood water, along with moisture that 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 a farmer must prioritize certain crop protection activities over others. 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 customized for use in assessments of the relative impacts of the top three yield-related concerns of farmers in the study area. As in the first case study, we formulated evidence-based specifications to derive this instance of the generic model to study the joint effect of pests & diseases, weeds, and available soil water on grain yield of rice and sorghum. To assess the effect of pests & diseases, weeds, available soil water and crop type on grain yield, the FBFS model was queried to simulate grain yields of rice and sorghum for FBFS settings in Kisumu County (Kenya) and Tigray region (Ethiopia). Each query was run 1000 times using the `mcSimulation` function (Luedeling et al., 2019). The derived model consists of three important levels of customization based on recommendations from the farmers. The first one, applied to the BN sub-model, consisted of selecting the two extreme states of the nodes of interest to derive a sequence of all possible combinations, which were later used as evidence to query the posterior distribution of the ‘farming constraint’ node (see **Figure 9** in the main paper material). The results of the BN queries were used to provide customized estimates of the ‘farming constraints’ node for the MC model, as described in section 4 and section on **Modelling interface** in the main paper. The second feature concerns the MC model where the `farming constraint` variable is assumed to directly affect crop growth by affecting the `biomass expansion factor` variable. The third one is that the average yield estimates used to feed the MC model were derived by the farmers who were asked to provide the ranges of values that are plausible on their farms. Scenarios were developed for the mid stage of crop development and the simulated grain yield data were analysed, accounting for uncertainty in estimates.

References

- Cullen, A., Frey, C., 1999. Probabilistic Techniques in Exposure Assessment. Plenum Press, New York.
- Delignette-Muller, M.L., Dutang, C., 2015. fitdistrplus : An R Package for Fitting Distributions. Journal of Statistical Software 64, 1–34. <https://doi.org/10.18637/jss.v064.i04>
- Luedeling, E., Goehring, L., Schiffrers, K., 2019. decisionSupport: Quantitative support of decision making under uncertainty. The Comprehensive R Archive Network.

Acknowledgements

This study was jointly funded by the Deutscher Akademischer Austauschdienst (German Academic Exchange Service; DAAD) and the World Agroforestry Centre (International Centre for Research in Agroforestry; ICRAF). We thank the farmers and experts who contributed in various ways during the 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. Thanks 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 drafted the models.