

Gebrekidan, Bisrat Haile; Heckeley, Thomas; Rasch, Sebastian

**Article — Published Version**

## Modeling intensification decisions in the Kilombero Valley floodplain: A Bayesian belief network approach

Agricultural Economics

**Provided in Cooperation with:**

John Wiley & Sons

*Suggested Citation:* Gebrekidan, Bisrat Haile; Heckeley, Thomas; Rasch, Sebastian (2022) : Modeling intensification decisions in the Kilombero Valley floodplain: A Bayesian belief network approach, Agricultural Economics, ISSN 1574-0862, Wiley, Hoboken, NJ, Vol. 54, Iss. 1, pp. 23-43, <https://doi.org/10.1111/agec.12740>

This Version is available at:

<https://hdl.handle.net/10419/287893>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



<http://creativecommons.org/licenses/by/4.0/>

# Modeling intensification decisions in the Kilombero Valley floodplain: A Bayesian belief network approach

**Bisrat Haile Gebrekidan** | **Thomas Heckeley** | **Sebastian Rasch**

Institute for Food and Resource  
Economics (ILR), University of Bonn,  
Nussallee 21, Bonn, 53115, Germany

## Correspondence

Bisrat Haile Gebrekidan, Institute for  
Food and Resource Economics (ILR),  
University of Bonn, Nussallee 21, Bonn,  
53115, Germany.  
Email: [bisrat.gebrekidan@ilr.uni-bonn.de](mailto:bisrat.gebrekidan@ilr.uni-bonn.de)

Data and the code used for this article is  
available in a github repository  
(<https://github.com/agpo-ilr-uni-bonn/CodeRepositoryForMS2021339>). A  
data appendix to replicate the main  
results is available in the online version of  
this article. Please note: Wiley-Blackwell  
is not responsible for the content or  
functionality of any [supporting  
information](#) supplied by the authors. Any  
queries (other than missing material)  
should be directed to the corresponding  
author for the article.

## Abstract

The Kilombero Valley floodplain in Tanzania is a major agricultural area. Government initiatives and projects supported by international funding have long sought to boost productivity. Due to increasing population pressure, smallholder farmers are forced to increase their output. Nevertheless, the level of intensification is still lower than what is considered necessary to increase production and support smallholder livelihoods significantly. This article aims to better understand farmers' intensification choices and their interdependent determinants. We propose a novel modeling approach for identifying determinants of intensification and their interrelationships by combining a Bayesian belief network (BBN), experimental design, and multivariate regression trees. Our approach complements existing lower-dimensional statistical models by considering uncertainty and providing an easily updatable model structure. The BBN is constructed and calibrated using data from a survey of 304 farm households. Our findings show how the data-driven BBN approach can be used to identify variables that influence farmers' decision to choose one technique over another. Furthermore, the most important drivers vary widely, depending on the intensification options being considered.

## KEYWORDS

agriculture, Bayesian belief network, intensification, Kilombero Valley, land use, regression trees, Tanzania

## JEL CLASSIFICATION

Q16, Q18, C11

## 1 | INTRODUCTION

The Tanzanian government places high priority on achieving food security while fostering sustainable development. The agricultural sector is the backbone of the country's economy and a major force behind rural development.

Around 80% of the workforce is employed in the sector, which also supports the livelihoods of over 70% of the population and contributes to approximately 95% of the national food requirements (Mwimo et al., 2016). This sector also contributes 28% of the gross domestic product and accounts for about 27% of export earnings (Milder et al.,

This is an open access article under the terms of the [Creative Commons Attribution](#) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *Agricultural Economics* published by Wiley Periodicals LLC on behalf of International Association of Agricultural Economists.

2013; Mwimo et al., 2016; United States Agency for International Development [USAID], 2019; World Food Program (WFP), 2019). However, with population growth outpacing production growth, food self-sufficiency declining, and malnutrition remaining high, subsistence agriculture still predominates (WFP, 2019).

The government pursues a policy of increasing domestic agricultural production, which is either driven by a switch to large-scale commercial farms or by an increase in smallholders' productivity as a result of opportunities and resources being made available to them (Coulson, 2015; URT, 2013). The government has sought to align these goals through different policy statements and national visions, including Kilimo Kwanza (Agriculture First), the Southern Agricultural Growth Corridor of Tanzania (SAG-COT), and Big Results Now (Coulson, 2015).

The wetland of the Kilombero Valley Floodplain (KVF) is one focus for the government in its effort to transform the country into a sustainable food basket. The low altitude plain with alluvial deposits has abundant natural resources, including fertile land, reliable water supply, and large pastures (Bamford et al., 2010; Nindi et al., 2014).

In the past, smallholders in the floodplain have benefited from an abundance of land to expand their agricultural production by cultivating new wetland areas and marginal lands (Figure 1) (Leemhuis et al., 2017; Msofe et al., 2019).

However, this type of land-use change is associated with various negative environmental consequences, such as loss of habitat and of above and underground biodiversity (Jones et al., 2012). Furthermore, increased immigration and population growth have pushed agricultural land expansion to its limit, and agricultural intensification has become the norm rather than the exception (Binswanger-Mkhize & Savastano, 2017; Kajisa, 2016; Otsuka & Place, 2013). As a result, with the support of a number of non-governmental organizations and private multinational corporations, the government has been promoting the use of optimized/high-quality inputs, adoption of new technologies or mechanization, and value-chain development as means of increasing productivity and closing yield gaps to generate sustainable and inclusive growth (Agra, 2016; Binswanger-Mkhize & Savastano, 2017; Otsuka & Larson, 2016). The current average rice and maize yield of farmers in the valley is 1.2 ton/ha, with more than half of farmers receiving less than 1 ton/ha, which is much lower than the 10–11 ton/ha of potential attainable yield under improved management practices and input-intensive rice and production systems in Tanzania (Nakano et al., 2018; Senthilkumar et al., 2018).

Although efforts are underway to accelerate and intensify production through increased adoption of improved technologies and optimal input use, smallholder farmers'

adoption of these options has been disappointing thus far (though gradually increasing) (Sheahan & Barrett, 2014). Feder et al. (1985) and Foster and Rosenzweig (2010) conducted extensive literature reviews on the adoption of different technologies in developing countries. A subset of this literature focuses on farmers' intensification decisions and the factors that influence relevant technology choices (Abay et al., 2016; Erenstein, 2006; Headey et al., 2014; Howley et al., 2012; Okike et al., 2001; Schelhas, 1996; Shriar, 2000, 2001; Wainaina et al., 2016). These sources indicate that the choices made by farmers are not simple reflexive responses to external drivers, but rather the result of complex decision-making. Farmers make decisions based on farm household socioeconomic characteristics, infrastructure, existing institutions, and the agro-ecological context (Wainaina et al., 2016). However, neither the current state of agricultural intensification nor the context- and locally specific factors influencing farmers' selection of specific intensification options are known for KVF (Milder et al., 2013; Milder et al., 2013; Nakano et al., 2016).

As noted by Vanlauwe et al. (2014) and Vanlauwe (2016), pathways toward intensification in Africa will necessitate a broad-based approach that considers a variety of options tailored to local agro-ecological conditions, crop choice and cropping patterns, farmer ability and willingness to invest, and specific institutional settings. Recent research has also revealed that farmers' perceived constraints and benefits play significant roles in their decision to pursue a specific option (Alomia-Hinojosa et al., 2018; Ntshangase et al., 2018; Yamano et al., 2015). Thus, it is integral to uncover factors that drive farmers' prioritization of options and their perceived constraints. The analysis of pathway-specific and locally relevant determinants of intensification in KVF could lead to the development of future policies that increase productivity through ecologically sustainable and pro-poor trajectories.

The article offers three novel contributions to the literature. First, we investigate how farm households decide on intensification when multiple pathways are available, highlighting the distinct factors influencing these decisions in a sensitive ecological landscape with high yield potential. Here, we focus on four land-saving intensification options: (1) using chemical fertilizers, (2) using improved seed, (3) using small-scale irrigation systems, and (4) increasing planting frequency. Second, we propose using a Bayesian belief network (BBN) as an analytical tool alternative to existing, typically lower-dimensional models (e.g., logit and probit models and decision trees). BBNs have the advantage of explicitly accounting for uncertainty and allowing for the incorporation of a diverse range of data types, including expert knowledge. They are structurally adaptable (Korb & Nicholson, 2010; Sun & Müller,

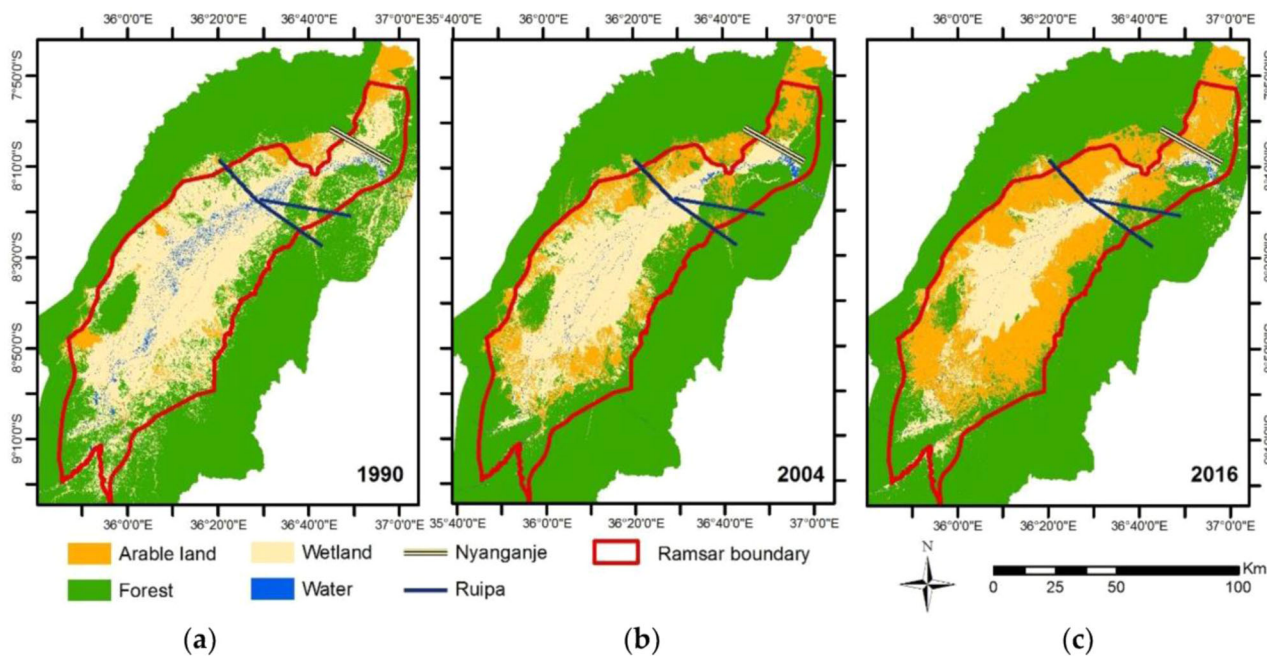


FIGURE 1 Land Use and land cover maps of the Kilombero floodplain for 1990 (a) 2004 (b) and 2016 (c) (Leemhuis et al., 2017, p. 8)

2013). Finally, we demonstrate how model-agnostic explanations (a combination of BBN, design of experiments, and regression trees) can be used to understand how various determinants influence intensification choices.

The remainder of this article is organized as follows. Section 2 provides a brief overview of our motivation for the methodological approach. Section 3 introduces the study site and data source. Section 4 describes the empirical model and variable selection in detail. Section 5 presents the results and discusses the intensification strategy options, and Section 6 concludes the article.

## 2 | MOTIVATION OF THE METHODOLOGICAL APPROACH

Different modeling tools have been proposed to understand individual decision-making regarding intensification alternatives. One strand of tools models the intensification decision as a binary choice problem involving adopting or rejecting a single intensification option. Prevalent approaches include probit models (Abay et al., 2016), logistic regression models (Erenstein, 2006; Okike et al., 2001; Perz, 2003), and decision trees (Gladwin, 1989), to name a few. For a review of modeling farmer adoption decision, see Besley and Case (1993). The second strand of tools examines the intensification option as one of several potential strategies using multivariate models, such as multinomial probit (Dorfman, 1996; Kassie et al., 2015;

Wainaina et al., 2016) and multinomial logit selection models (Kassie et al., 2018; Khonje et al., 2018; Teklewold et al., 2013). We propose a BBN as an alternative modeling technique for multivariate data. BBN, also known as Bayesian net, causal probabilistic network, Bayesian network, or simply belief network, is a probabilistic graphical modeling tool that permits the representation of knowledge and facilitates reasoning under uncertainty (Kjaerulff & Madsen, 2012; Korb & Nicholson, 2010; Pearl, 2009). Similar to other graphical models, the nodes represent stochastic variables, and the arcs direct dependencies based on process understanding, statistical, or other types of associations between the linked variables (Chen & Pollino, 2012). Formally, the Bayesian network is a directed acyclic graph (DAG) that defines a factorization of a joint probability distribution over variables, where the directed links of the DAG provide the factorization. Specifically, for a DAG  $G(V, E)$ , where  $G$  denotes graph,  $V$  denotes a set of nodes, and  $E$  is a set of directed links (or edges) between pairs of the nodes, a joint probability distribution,  $P(X_V)$ , over the set of variables  $X_v$   $v \in V$  can be factorized as

$$P(X_V) = \prod_{v \in V} P(X_v | X_{pa(v)})$$

where  $X_{pa(v)}$  is a set of parent nodes for variable  $X$  at node  $v$  (Kjaerulff & Madsen, 2012).

In addition to jointly considering multiple pathways of intensification choices, our methodological choice is driven by the following benefits of BBNs:

- a. The relevance of random events and the complexity of farmers' decision-making processes favor an approach that explicitly addresses uncertainty. The Bayesian approach to uncertainty ensures that the system remains consistent, which also provides a direct way to apply the model to data (Koski & Noble, 2011). Because BBNs are joint probability distributions, uncertainty is reflected in the model's output. In contrast to deterministic models, the probabilistic representation of knowledge in BBN prevents overconfidence in the potency of responses obtained by simulating changes in one or more variables of interest (Uusitalo, 2007).
- b. BBN provides generality and formalism of displaying relationships clearly and intuitively, in contrast to other "black-box" models (Daly et al., 2011; Margaritis, 2003). They can include quantitative data and the qualitative opinions and viewpoints of experts and other key players (Daly et al., 2011; Sun & Müller, 2013; Uusitalo, 2007).
- c. BBNs are easily updated as more data become available.
- d. Because Bayesian networks with discrete variables are non-parametric models, they do not require assumptions about the distribution of the variables. Additionally, probabilistic methods can infer nonlinear functions (Blattberg et al., 2008, p. 486).
- e. Another significant advantage of Bayesian networks over other models is their capacity for both forward propagation (prediction) and backward (diagnostic) inferences. For instance, by enabling models to evaluate likely outcomes based on assumed states of various factors, forward propagation can address prediction questions, whereas backward inference can address profile questions, such as "What is the probability that an irrigator is located in the center of the floodplain?" or "What is the probability that a non-intensifier is a young household head?" (Blattberg et al., 2008; Uusitalo, 2007)

Despite the advantages of BBN as a modeling tool, some difficulties are inherent in its development and application. The primary disadvantage of BBNs is the lack of a widely accepted mechanism for deriving network structure from data. Despite efforts, most models are currently built using domain knowledge (expert advice) or a combination of machine learning and empirical data (Kyrimi et al., 2021). Another significant limitation of BBN modeling is its reliance on discrete variables. They are also computationally expensive. Because inference in a Bayesian network necessitates summarizing over an exponential number of terms, this task requires massive processing power (Blattberg et al., 2008).

Bayesian belief networks are widely used in diverse domains, including medicine, environmental modeling, natural resource management, and forecasting (Daly et al., 2011; Korb & Nicholson, 2010; Uusitalo, 2007). However, only a few BBN applications to farm management are known; see Drury et al. (2017) for a review of BBN applications in agriculture. Cain (2001) employed a BBN to investigate crop yield determinants, and Prishchepov et al. (2019) used a BBN to examine wheat yield determinants in Siberia. Sun and Müller (2013) choose a BBN to simulate land conversion patterns by modeling the binary choice of participating in a scheme with payments for ecosystem services and combining it with opinion dynamics in an agent-based modeling framework. Meanwhile, Frayer et al. (2014) developed a BBN to analyze the drivers of the decision to plant trees on former cropland in a similar study. Aalders (2008) and Celio et al. (2014) built a BBN to incorporate farmers' preferences for various land-use options. Meanwhile, Rasmussen et al. (2013) used Farmers Agricultural Data Network data to create a large-scale BBN tool for risk management in EU agriculture. Pope and Gimblett (2017) investigated the various ranching strategies farmers choose under varying environmental conditions using BBN and agent-based modeling. Moreover, Ticehurst et al. (2011) investigated landholder adoption of conservation practices decisions (fencing of native bushland) in the Wimmera region of southeast Australia using BBN. To our knowledge, no other study has used a BBN to model farmers' adoption of multiple intensification options.

### 3 | CONTEXT AND DATA

#### 3.1 | Study site

This research was conducted in KVF, Tanzania. The low-elevation plain with alluvial deposits possesses a productive natural resource base with fertile land, dependable water availability, and extensive pastures (Bamford et al., 2010; Nindi et al., 2014). It is one of the four principal subbasins of the Rufiji River Basin, located in southern Tanzania's Ulanga and Kilombero districts (Figure 2). It consists of several rivers and seasonally flooded marshes and swamps (Dinesen, 2016). The seasonal variation in water levels is significant. During the wet season, the plains are completely submerged, whereas the water fully retracts to rivers, river margins, and areas with permanent swamps and water bodies (Kato, 2007; Ntongani et al., 2014).

The KVF contains the Kilombero Game Controlled Area, which encompasses approximately 7000 km<sup>2</sup>, and

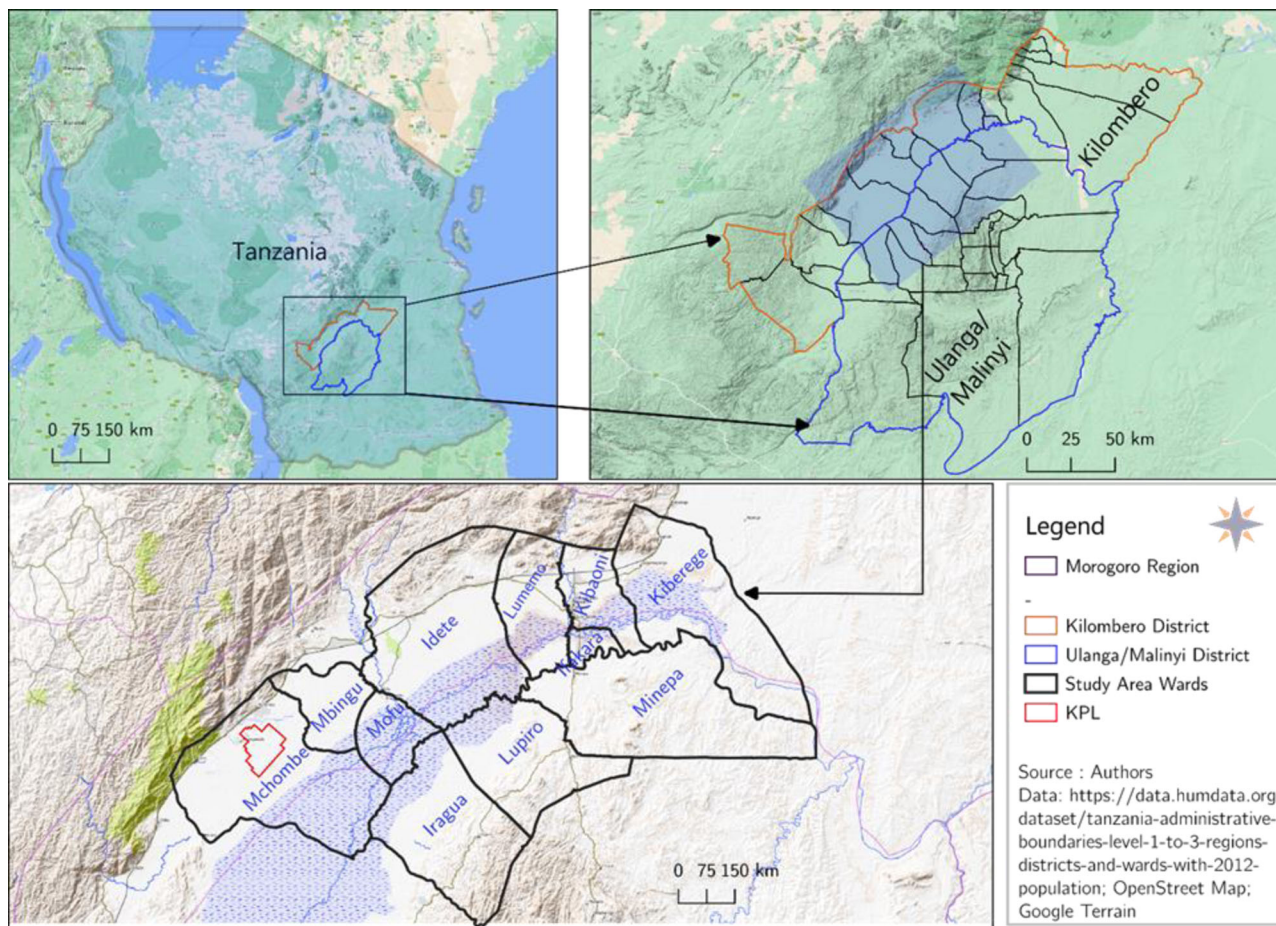


FIGURE 2 Location of the study site

the Kilombero Valley Ramsar site, which comprises 7976 km<sup>2</sup> (Dinesen, 2016; Nindi et al., 2014). The Kilombero Valley, one of Africa's most extensive wetlands, has a long history of agricultural production (Kato, 2007; McCartney et al., 2010). The floodplain provides crops, fish, drinking water, forest products, and fuelwood to more than 500,000 people (Government of Tanzania, 2013), thereby contributing significantly to their standard of living (Mombo et al., 2011). This area has witnessed rapid growth in agricultural land utilization (Jones et al., 2012). Immigration into the valley has increased dramatically due to the perceived availability of high-quality and cheap farmland. Moreover, conflicts between pastoralists and farmers over land use are pervasive and persistent, resulting in injury and litigation disputes (Dinesen, 2016; MALF, 2015; Nindi et al., 2014).

### 3.2 | Data

The primary data source is a household survey conducted in 21 villages across the Ulanga and Kilombero districts of the Kilombero Valley. In total, 304 farm households were

surveyed. Data were gathered using a household questionnaire, which also included demographics data, information on land use (crop selection), labor use, physical crop output quantities, specific input use at the plot level, and the household's embeddedness in social networks and institutions (the full questionnaire and descriptive statistics of the respondents are attached with supplementary material). Six well-trained (university graduates) enumerators with a working knowledge of the farming environment and the regional language administered the questionnaire during a 2-h in-person interview. The interview's questions were initially written in English and then translated into Swahili by field assistants. A pre-test survey was also carried out to ensure the enumerators could properly administer the questionnaire and the farmers understood the questions. Several questions and possible answers have been rewritten in light of the pre-test to improve participant comprehension. Following a final clean-up, we obtained 297 observations with complete data and seven observations (2.3%) with missing data in one or more variables.

The household selection was based on a multi-stage sampling strategy. First, 11 wards were purposively selected

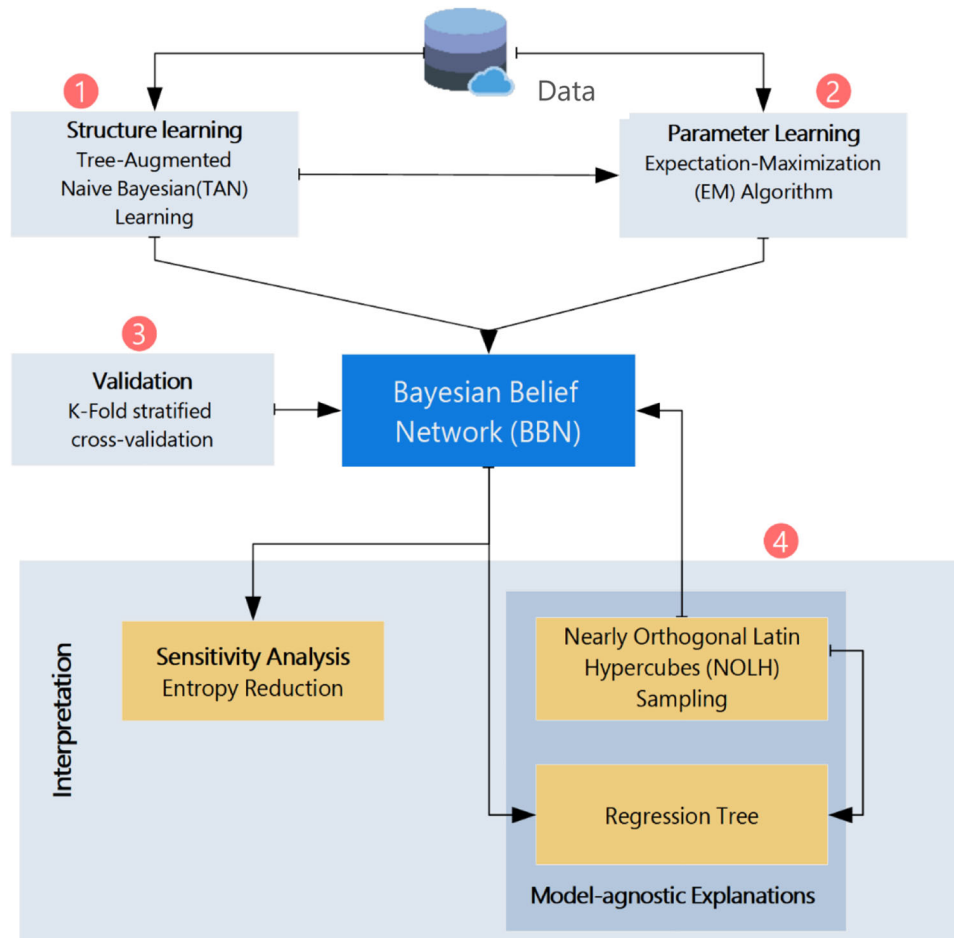


FIGURE 3 Workflow of our modeling approach

based on the presence of floodplain farming. In the second stage, 21 villages were chosen at random based on population size within the wards. In the final stage, households were chosen at random from a list provided by village leaders. To estimate the study area's boundaries and total population size, we employed a GIS approach incorporating the land use map from GLC30 (Jun et al., 2014), the administrative boundary, and 2012 census data from the Tanzania statistics office. We used a digital elevation model at 90-m resolution from the shuttle radar topography mission to drive the biophysical characteristics of farmer plots (slope, elevation, roughness) (Jarvis et al., 2008).

#### 4 | METHODOLOGY TO EMPIRICALLY SPECIFY, VALIDATE, AND INTERPRET THE BBN

This section describes our methodology for generating, validating, and interpreting the results. As shown in Figure 3, our empirical modeling strategy comprises four steps. The

specification or learning of a BBN includes (1) structure learning and (2) parameter learning, which correspond to model selection and parameter estimation in conventional statistical models, respectively (Koller & Friedman, 2009; Nagarajan et al., 2013). The learning is then followed by (3) k-fold stratified cross-validation to evaluate the model's performance and by (4) interpretations of the final BBN via sensitivity analysis and model-agnostic explanations. In the subsequent subsections, we will highlight each step in greater depth.

##### 4.1 | Structure learning

In general, there are two methods for creating a BBN's structure. It can be learned through knowledge of engineering from experts, literature, and theory (Frank, 2015). Experts' opinions on the correlations between random variables are helpful in the learning process. Structure learning can use experts' opinions to learn more accurate network architectures than can be learned from data alone (Amirkhani et al., 2017).

Alternatively, it can be learned from data. As further described, we use a data-based approach in this study supplemented with theory and plausibility.

Finding dependencies between variables that produce distributions in the probability space that are as close as possible to the observed data is a key step in the empirical method of learning a BBN.<sup>1</sup> The following are two general classes of algorithms for learning a Bayesian network's structure from data:

#### 4.1.1 Constraint-based structure learning

These approaches view a Bayesian structure learning network as a representation of interdependencies. The approach selects conditional dependence and independence based on statistical tests (such as chi-squared or mutual information) and uses these relationships as constraints to construct a BBN (Koller & Friedman, 2009; Neapolitan, 2010). These algorithms include inductive causation, grow-shrink, and incremental association.

#### 4.1.2 Score-based structure learning

This optimization-based search treats a Bayesian network as though it were a statistical model that needs to be specified. It generates potential Bayesian network candidates, scores each candidate individually, and then returns the candidate with the highest score (Kjaerulff & Madsen, 2012; Nielsen & Jensen, 2009).

We used a score-based structure learning variant called a tree-augmented naive (TAN) Bayesian network for this study (Friedman et al., 1997). It relaxes the assumption of attribute-independence of Naive Bayes by imposing constraints on the network structure and selects the tree that maximizes the likelihood conditional on the training data (Koller & Friedman, 2009; NorsysSoftwareCorp, 2016; Zheng & Webb, 2010). According to Friedman et al. (1997), using TAN to learn a network's structure achieves a good trade-off between estimation quality and computational complexity. We iteratively modify the network structure after it has been created using TAN, as suggested by Frayer et al. (2014), Neapolitan (2010), Prishchepov et al. (2019), Sun and Müller (2013) and Ragno et al. (2022). During this process, some links' direction is reversed, and new links are added. For example, the relationship between household age, crop choice, and topographic wetness index is reversed based on domain knowledge.

<sup>1</sup> For a detailed explanation of algorithms learning the Bayesian network structure from data see Koller and Friedman (2009) and Nielsen and Jensen (2009).

To capture the dependency between the two nodes, we added a link between the share of nonfarm income and per capita income. The structure learning process is then repeated with the newly constrained link. However, modifying the model learned through TAN based on theoretical knowledge may reduce the model's accuracy and performance. Appendix B compares our final model's model accuracy to unmodified TAN. Our post-processing indeed increases the average error rate of the TAN model by 4%. Hence, the final model we chose involved a trade-off between plausibility and prediction accuracy. Netica (5.4) was used to create the structure (NorsysSoftwareCorp, 2016)<sup>2</sup>.

## 4.2 | Parameter learning

Several alternative ways of parameter learning, that is, generating estimates for the conditional probabilities (CPTs), are available for BBN structure learning. A maximum likelihood procedure was used to derive the probabilities from the survey data. We opted for learning the CPTs from data to reduce the number of conditional probability entries required from experts or the literature and to learn objective rather than subjective probabilities. Some algorithms can solve the underlying maximum likelihood problem (e.g., count learning, expectation-maximization (EM), and gradient descent (GD); Frank, 2015; NorsysSoftwareCorp, 2016). Our final network is based on EM learning, as proposed by Uusetello (2007). EM has an advantage over learning parameters because it converges more robustly and provides slightly better accuracy, despite the number of missing observations being minimal (2.3%). Appendix B presents model accuracy using EM and GD for robustness testing.

## 4.3 | Validation of the BBN

Validating the BBN is essential for ensuring the model's quality. There are several quantitative and qualitative methods for validating the constructed BBN. Qualitative ones check the validity using an expert opinion (Celio et al., 2014; Frank, 2015). Meanwhile, quantitative

<sup>2</sup> Netica provides numerous simplifying tasks for the modeler, including a high visual capability to display the network and advanced algorithms to learn the structure and parameters of the network. Using the Java API for the construction of BBNs provides an advantage in terms of transparency, reproducibility, and easy integration with other modelling tools of interest as it resulted in Java source code. The full documentation and Java code, Python code for regression tree, and the data can be found as [supplementary material](https://github.com/agpo-ilor-uni-bonn/CodeRepositoryForMS2021339) a GitHub link <https://github.com/agpo-ilor-uni-bonn/CodeRepositoryForMS2021339>



validation employs a test data set not used in parameter learning to evaluate the accuracy of the target variable's predictions.

Our BBN is quantitatively validated using the *n*-fold cross-validation method. It is a resampling technique that employs multiple repetitions of data, each with a distinct subset of observations, to test and train a model. Cross-validation assesses a model's ability to predict data not used during estimation. This aids in identifying problems such as overfitting and bias and provides insight into how the model will generalize to unknown data points (Hastie et al., 2009).

Cross-validation in a single round involves dividing a data sample into complementary subsets, performing analysis on one subset (called the training set), and confirming the analysis on the other subset (called the validation set or test set). To reduce variability, we conducted multiple rounds of cross-validation with distinct subsets and pooled the validation results (e.g., averaged) across rounds to determine the model's accuracy.

We implement a fivefold cross-validation strategy (Hastie et al., 2009, p. 243) by partitioning our data into five disjoint subsets, followed by an iterative validation of the parameters. The error rate is then calculated using the confusion matrix. The error rate indicates the frequency with which the network predicted an incorrect value. To account for the unbalanced nature of our target node and ensure that all states are represented equally in the split, we employ stratified cross-validation (Kuhn, 2008).

## 4.4 | Model interpretation

The final step is to interpret our model (i.e., explain the BBN) after its construction and parameter validation. To determine the impact of various determinants on choices of intensification, we employ two distinct approaches: (1) sensitivity analysis and (2) model-agnostic explanation. The first identifies the most important variables for determining the probability distribution across intensification options (Rohmer & Gehl, 2020; Sun & Müller, 2013), whereas the second identifies the variables' specific contribution to predicting each option.

### 4.4.1 | Sensitivity analysis

We use sensitivity analysis to measure changes in the overall probability distribution of the target node with changes in states of one of the input nodes (Pollino et al., 2007; Rohmer & Gehl, 2020). The entropy reduction (mutual information) method is used in this case to evaluate the sensitivity of the BBN model's output to

changes in a specific input parameter because the input nodes necessary for the sensitivity analysis contain discrete values. The entropy reduction method calculates the anticipated decrease in entropy at the target node (i.e., the increase in information) due to findings at another child node. It is calculated as follows (Marcot et al., 2006; NorsysSoftwareCorp, 2016; Pearl, 1988):

$$I = H(Q) - H(Q|F) = \sum_q \sum_f \frac{P(q, f) \log(P(q, f))}{P(q)P(f)}$$

where  $H(Q)$  and  $H(Q/F)$  are the entropy of node  $Q$  before and after any new findings in node  $F$ , respectively.  $q$  is the state of the query variable, and  $f$  is the state of the varying variable.

### 4.4.2 | Model-agnostic explanation

Although the sensitivity analysis provides insight into the importance of influencing factors in explaining variations (via entropy reduction) in the overall probability distribution of intensification strategies, it does not tell us how the variables included in our model influence the probability of each strategy. To deliver such information, we conduct a model-agnostic interpretation (Molnar, 2019, p. 110), combining a BBN simulation based on the design of experiments with a meta-modeling approach. Our approach is based on a widely used post-hoc model-agnostic technique known as local interpretable model-agnostic explanation (LIME), which explains the predictions of any classifier or predictor by fitting a less complex model locally around the prediction (Molnar, 2019; Ribeiro et al., 2016). LIME uses perturbed samples of a given instance in the feature space and observes the effect of these new sample points on the classifier's output (Molnar, 2019).

Four steps comprise our model-agnostic approach, (i) The perturbations of the network's entry values  $C$  (i.e., the determinants) are generated by drawing from probability distributions that take into account the correlation between the features using nearly orthogonal Latin hypercube (NOLH) sampling (Sanchez, 2005), (ii) The query probability of interest  $P$  derived from the inferences is estimated using the BBN, and the probabilities of each strategy for each sample point are recorded, (iii) The link between  $C$  and  $P$  are established using a regression tree. The incorporation of nonlinear interactions, the minimal assumptions about the data structure required, their robustness to outliers, and the implicit handling of variable selection are all advantages of regression tree (Couatts & Yokomizo, 2014; Kuhn & Johnson, 2013). Furthermore, it serves as an analogy for natural rule induction from results.

The regression trees were implemented using Scikit-learn (machine learning tool in Python; Pedregosa et al., 2011), and (iv) The regression tree results are interpreted using feature importance (the most important predictor) and tree rule induction.

## 5 | VARIABLE SELECTION

In general, the factors that influence the choice of an intensification strategy fall into three broad categories: (1) socioeconomic characteristics of farm households, (2) infrastructure and institutional factors, and (3) the farm's agro-ecological context (Wainaina et al., 2016). Numerous studies found that household resource endowments in terms of labor, land, capital, risk behavior, and social capital influence the selection of a specific strategy (Erenstein, 2006; Feder et al., 1985; Ghadim et al., 2005; Okike et al., 2001). For example, increasing cropping frequency is frequently constrained by labor availability. Moreover, human capital acquired through education, training, and experience affects the decisions of farmers (Kijima, 2016; Wainaina et al., 2016). Furthermore, numerous empirical studies demonstrate that infrastructure and institutions, such as distance to the nearest market or access to credit and agricultural extension, are significant (Feder et al., 1985; Kassie et al., 2015, 2018; Teklewold et al., 2013). Finally, biophysical characteristics of the farm plots, including slope, hydrological regime, and soil characteristics, influence the choice of one strategy over another (Khonje et al., 2018; Nkonya et al., 1997; Sirén, 2007).

To avoid unnecessary model complexity, we employ a data mining technique to select the most significant variables for explaining variation in intensification selection. In the first step, we encoded a farmer as either an intensifier (if the farmer adopted at least one option) or a non-intensifier using a binary variable. Scikit-learn (Python's machine learning library) is then utilized to execute a random forest algorithm between the binary choice variable and a variety of explanatory variables (Pedregosa et al., 2011). Recently, random forest algorithms have gained popularity in variable selection. See Genauer et al. (2010), Rogers and Gunn (2005), Rogers and Gunn (2006) and Sandri and Zuccolotto (2006) for information regarding their application and use in variable selection.

Our variable of interest (target node), *intensification*, is treated as a discrete node containing four intensification options (use of improved seed variety, small-scale irrigation, fertilizer application, and multi-season farming) and all possible combinations of these strategy bundles. Also included is a state that captures the absence of an intensification strategy. Because only five strategy sets were observed in our data set, "others" represent the unob-

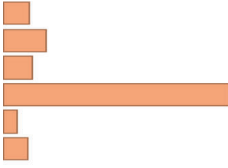

served strategy combinations. Here, we can take advantage of BBN's capability to update the conditional probabilities when new data are collected.

We used per capita income as a surrogate for a resource endowment and availability of capital. It includes income from agriculture (farming and fishing), off-farm activities, land rental, and brickmaking. To represent the farming practices used by a particular household, we also include the farmer type. The farmer type variable is a typology constructed through a combination of principal component analysis and hierarchical clustering to stratify farmers into homogeneous clusters according to their livelihood and land use (Gebrekidan et al., 2020). We generated a topographic wetness index (TWI) using a digital elevation model of our study area (slope and upslope contributing area) to capture the quality and hydrological characteristics of the farm. According to Sørensen et al. (2006), the index indicates the relative wetness within the catchment and is highly correlated with soil moisture and ground-level water. It is especially helpful in situations with limited data availability and high spatial variations because it only uses one data source (i.e., digital elevation model; Kopecký et al., 2021; Sørensen et al., 2006). Crop selection and anticipated income will be impacted by expected prices. Based on the household's proximity to the market, we included the prices paid for rice and maize as expected prices for the two crops. Due to a lack of past price data, we assume farmers expect what they received during planting. Our distance variable expresses the distance in kilometers between the farm and the closest large market. When determining distance from a farm rather than a homestead, access to and the cost of transportation from the farm to either the farmers homestead or the market are taken into consideration. While homesteads are clustered around the single main road, farms that are in the bottom valley are quite far from homesteads.

Although BBNs can typically handle continuous nodes, the handling of continuous variables is currently constrained by methods and software applications. Hence, we followed a common practice of discretizing all continuous variables into smaller classes (Frayer et al., 2014; Sun & Müller, 2013). We use a heuristic method called "equal interval" to partition our continuous variables into groups of  $K$  equal lengths or width (Clarke & Barton, 2000; Nojavan et al., 2017). However, we cannot rule out information loss, and the resulting discretized variable closely resembles the continuous variable's distribution, which may also affect the model's accuracy.

Table 1 presents the descriptive statistics of our target node and 15 evidence nodes included in the final network. Regarding the target node (choice of intensification), approximately 62% of our sample households did not intensify their production, whereas 38% chose one or

TABLE 1 Descriptive statistics for the variables included in the final network

Variable	Unit	Mean (SD)	Distribution
Intensification options		21 (7.1%)	
1. Apply Fertilizer		35 (11.8%)	
2. Apply Improved Seed		24 (8.1%)	
3. Crop Multiple Times		186 (62.6%)	
4. None		11 (3.7%)	
5. Use Irrigation		20 (6.7%)	
6. Use Irrigation + Fertilizer			
Age of the Household Head	Year	46.41 (12.82)	
Household Size	Number	5.13 (2.18)	
Farm Size	Hectares	2.64 (2.83)	
Share of Hired Labor	Percent	36.41 (32.98)	
Commercialization Index	Percent	46.87 (24.99)	
Total Labor	Man-Days/ per hectares	321.94 (343.58)	
Share of Nonfarm Income	Percent	8.1 (18.37)	
Topographic Wetness Index	Index	20.73 (4.49)	
Distance to the nearest big Market	Km	22.1 (16.44)	
Maize Price	Tanzanian Shilling (Tsh)/kg	382.01 (425.44)	
Rice Price	Tanzanian Shilling (Tsh)/kg	1232.38 (260.49)	
Income	000'Tanzanian Shilling (Tsh)	501.58 (1048.67)	
Farmer Type		21 (7.1%)	
1. Agro-Pastoralist		81 (27.3%)	
2. Diversifier		195 (65.7%)	
3. Mono-Crop Rice Producers			
Credit Access		199 (67.0%)	
1. No		98 (33.0%)	
2. Yes			
Crop Choice		5 (1.7%)	
1. Maize		148 (49.8%)	
2. Rice		100 (33.7%)	
3. Rice + Maize		20 (6.7%)	
4. Rice + Maize + Vegetables		22 (7.4%)	
5. Vegetables + Rice			

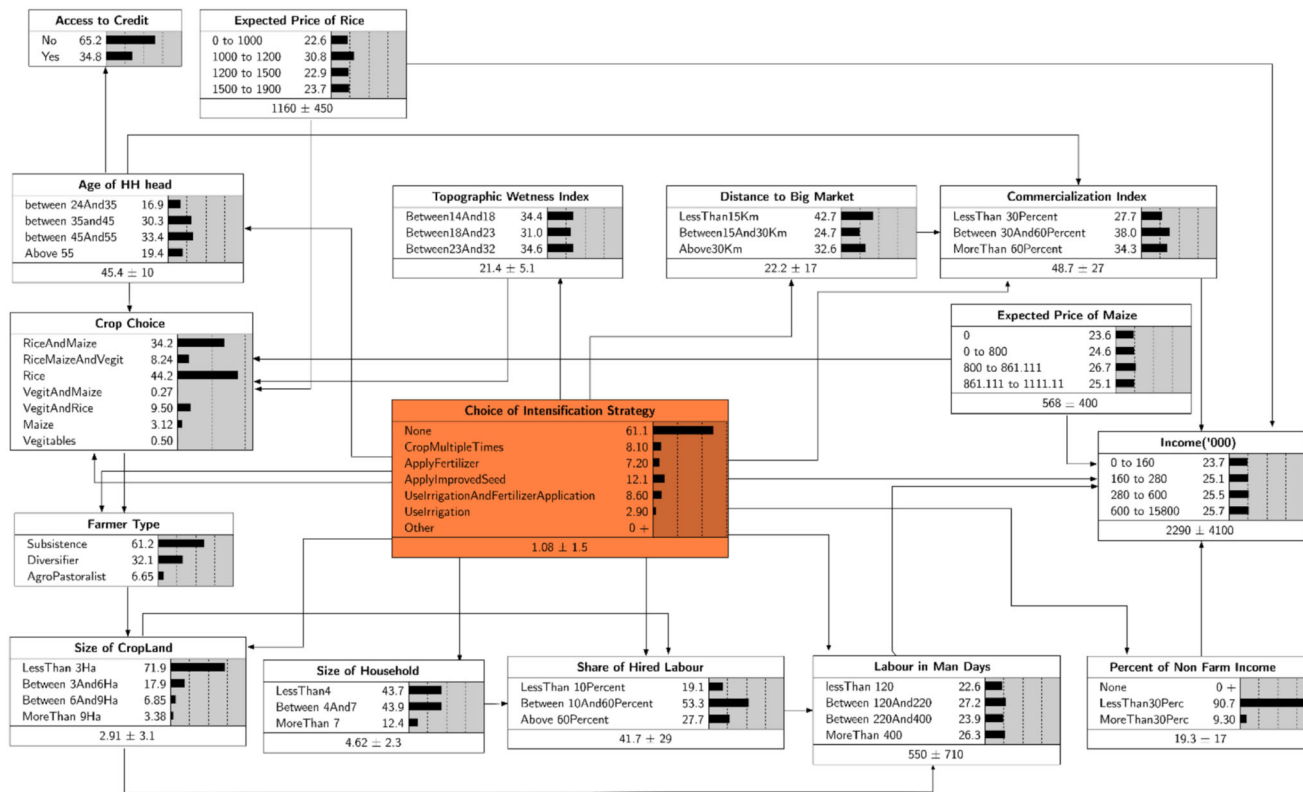


FIGURE 4 A Bayesian belief network of intensification decision in KVF.

Note: To reduce the complexity of the CPT and easy learning of the parameters, the links leave the target node rather than pointing to it and also represent the application of the BBN for backward propagation

more intensification options. Twelve percent of farmers use improved seed varieties, 8% plant in short and long rainy seasons, 7% use chemical fertilizers, 6.73% use irrigation and chemical fertilizer together, and only 3.7% use irrigation.

## 6 | RESULTS AND DISCUSSION

### 6.1 | Factors affecting the choice of intensification strategy

Figure 4 depicts the final learned Bayesian belief network, highlighting key relationships for determining KVF intensification decisions. All variables are associated with the choice of intensification strategy by design (TAN structure learning algorithms). Some of the input nodes are also correlated. For example, the age of the household head is correlated with farmers’ commercialization index, credit availability, and crop selection. The proportion of hired labor is also related to cropland and household size. Moreover, topographic wetness index of the plot influences the farmer’s crop choices.

Furthermore, the distance from the nearest large market has a direct impact on the expected prices of rice and

maize, and the amount of output sold to the market. The size of cultivated land is directly related to the proportion of hired labor and total labor use in person-days. Another link exists between per capita income, market participation, the proportion of hired labor, and access to nonfarm income.

The BBN also shows the posterior probabilities learned from the data using the EM algorithm. The probabilities match the data, demonstrating that the parameter learning algorithm can learn the underlying joint distribution and conditional probability tables, given the network structure. However, we perform a fivefold stratified cross-validation to fully assess the quality of both structure and parameter learning (see model validation section above). The resulting confusion matrix has an average error rate of 45% predicting the choice of intensification strategy. Error rates are generally relative measures that are interpreted concerning the cost associated with the model’s objective and decision. We compared our model to Naive Bayes, another classifier, to determine whether the high error rate was due to our modeling choices (Appendixes B and C). Even though the error rates in prediction are not significantly different between Naive Bayes and the BBN, the BBN provides more probabilistic links between the determinants and the choices, allowing policy implications to

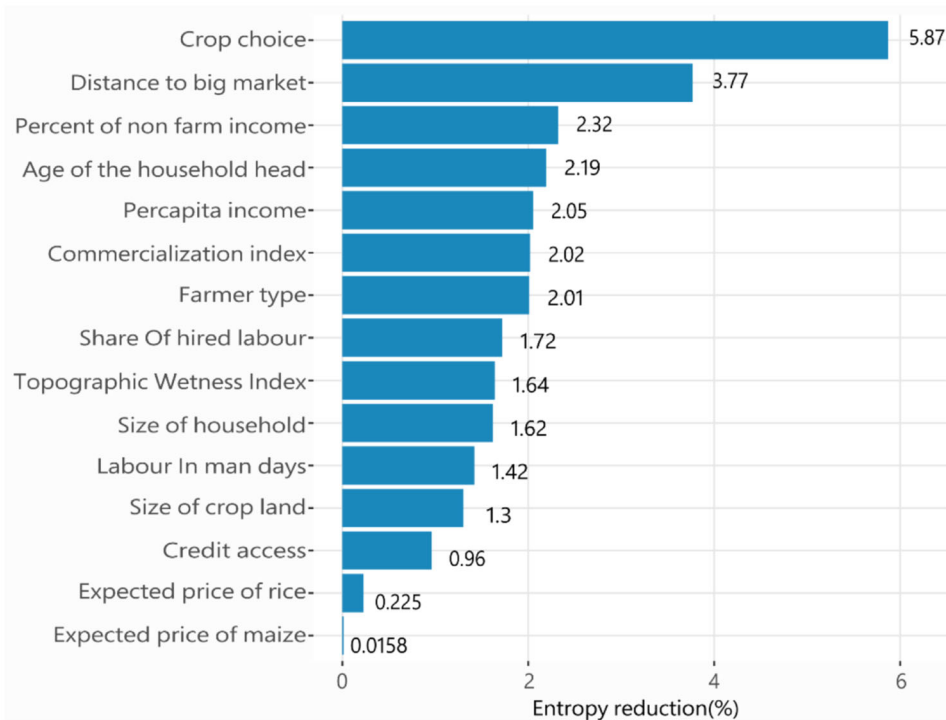


FIGURE 5 Sensitivity analysis of intensification choice

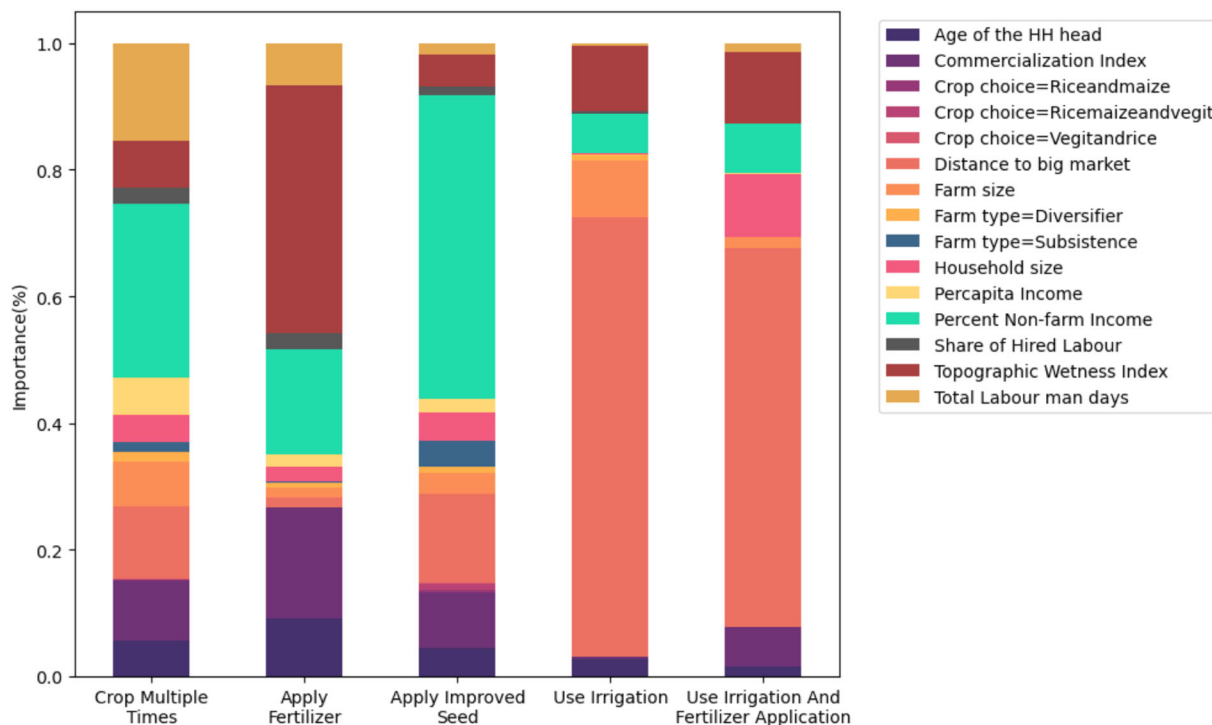
be deduced. To reduce error rates even further, one could obtain additional data to ensure that the model has enough observations to learn all relevant patterns, compare various discretization methods and levels, and finally, obtain prior probabilities from experts and stakeholders.

The results of the sensitivity analysis are shown in Figure 5. The graph compares each variable's contribution to the total expected entropy reduction of our target node, that is, choice of intensification strategy. A closer inspection of Figure 5 indicates that crop choices strongly influence the choice of intensification strategy with a 5.87% reduction in entropy. This suggests that farmers who cultivate particular crops are more likely to employ one type of intensification over another. Given that rice is the predominant crop in the region, crop selection variations primarily involve rice, maize, and vegetables. Additionally, distance from the farm to the closest market and commercialization substantially impact the selection of intensification strategy (with 3.8% and 2.02% entropy reduction, respectively). This is consistent with the findings of Erenstein (2006a), who discovered that market access affects both access to key input and output markets, thereby significantly affecting intensification. The outcome also indicates that nonfarm income has a significant impact. The importance of off-farm income in the adoption decision is widely recognized. Off-farm income

is a substitute for borrowed capital in rural economies with absent or dysfunctional credit markets (Reardon et al., 2007).

The farmer characteristics are of varying importance: Age of the household and per capita income significantly impacts the choice of intensification measure, whereas the size of the household and the amount of land available have a negligible impact. Access to more information or variances in risk aversion may influence the choice of intensification as a function of age (Kariyasa & Dewi, 2013; Mwangi & Kariuki, 2015). In addition, the per capita income has a strong influence on the strategy farmers prioritize over others. As a proxy for farmers' endowment, per capita income indicates their capacity to invest additional resources and assume the risks associated with adopting new options (Xie & Huang, 2021). Although the availability of family and hired labor is regarded as a crucial determinant of intensification choice (Lee, 2005; Wainaina et al., 2016), our sensitivity analysis indicates only a moderate impact.

Farmers' intensification is also influenced by the ecological context of the farm, as indicated by the topographic wetness index of the plots. One possible explanation is that the ecological context determines whether or not a particular option is applicable to a given farm, as well as how much it costs to implement and what other options are



**FIGURE 6** Variable importance for intensification decisions. Note: The data underlying the regression trees are posterior probabilities from our BBN using 257 sample points generated through NOLH design that captured the structure of the conditional distribution between determinants

already in place on that farm (e.g., farms located within the floodplain may not need to irrigate their plot (Pronti et al., 2020; Yesuf & Köhlin, 2008).

## 6.2 | Factors affecting a specific intensification strategy

The meta-modeling analysis reveals the variables influencing a selection of specific intensification strategy. The main finding is that although all options have a similar set of variables that affect their probabilities, differences exist in the magnitude and order of those effects. A variable importance plot for the regression trees for each intensification strategy can be seen in Figure 6. The variable importance is determined based on the total (normalized) reduction of the residual sum of squares caused by a particular variable. A high value denotes an important predictor (James et al., 2013).

The findings from the regression tree suggest that factors such as total labor available throughout the year, the commercialization index, the topographic wetness index, income, and distance to the central market are the best predictors of whether a farmer will decide to plant multiple crops. Having enough labor supply is one of the

pre-requisites for cropping multiple times, which suggests a potential limiting factor in the KVF given that household members may alternatively travel to the city for work during the dry season. The decision to plant is also influenced by market accessibility and availability. Access to markets encourages farmers to double crop rice–rice, rice–maize, and rice–vegetables, thereby producing excess quantities for sale on the neighborhood market.

The use of improved seeds is best predicted by the share of nonfarm income, age, household size, distance to the market, and farm size. These findings are consistent with previous adoption studies. For example, age is one of the human capital characteristics frequently associated with the adoption of improved varieties (Kafle, 2010). Similarly, Salasya et al. (2007) discovered a strong correlation between market distance and adoption of improved variety (stress-tolerant maize hybrid).

Meanwhile, market proximity has the largest influence on the likelihood of using irrigation and fertilizer. Farmers with better market access may obtain a better quality and more sustainable supply of fertilizers (Kafle, 2010), allowing them to cultivate vegetables and other high-value crops sold on the market to supplement their income. Furthermore, farm size, nonfarm income share, and topographic wetness index all influence the decision to use irrigation.

TABLE 2 Factors discriminating highest and lowest probability of choosing a specific intensification option

<b>Cropping multiple times</b>	
High Probability = .338	Low Probability = .107
Topographic wetness index > 18.9	Topographic wetness index > 18.9
Commercialization Index < 66.6%	Commercialization Index < 66.6%
Distance from market > 44 km	Distance from market < = 44 km
Commercialization Index < 36.14%	Total labor (man days) < = 615
	Income ('000 Tsh) > 228
<b>Improved seed</b>	
High Probability = .282	Low Probability = .051
Household size < 5	Household size > 4
Distance to the market > = 23 km	Age of household head < = 51 years
Farm size > 10 ha	Commercialization Index < 72 %
	Farm size > 9 ha
<b>Small-scale irrigation</b>	
High probability = .058	Low probability = .008
Labor in man-days > 110	Labor in man-days > 110
Farmer type! = Mono-crop rice producer	Farmer type! = Mono-crop rice producer
Topographic wetness index > 28.78	Topographic wetness index < 28.78
	Share of non-farm income < 30 %
	Household size > 5
<b>Fertilizer application</b>	
High probability = .276	Low probability = .068068
Farm size > 5 ha	Farm size < = 5 ha
Topographic wetness index < 28	Distance to the market < = 35 km
<b>Irrigation and fertilizer</b>	
High probability = .223	Low probability = .0228
Share of non -farm income > 38.13%	Share of non -farm income < = 38.13%
Farm size > 7.4 ha	Topographic wetness index < 28
Distance to the market > 22 km	Farmer type ≠ diversifier
	Share of nonfarm income > 30 %

Commercialization, or growing for profit rather than subsistence, is another important factor influencing whether a farmer will apply fertilizer or irrigate his or her crops.

One of the primary benefits of our modeling approach is that it allows us to see how specific attributes, critical thresholds, and nonlinear interactions influence the likelihood of selecting a specific strategy. Table 2 shows the variable value ranges that result in the highest and lowest probability of selecting a specific option (The complete regression trees are presented in Appendix A).

The likelihood of multiple harvests is greatest when the TWI is greater than 18.9, the commercialization index is between 36.14% and 66.6%, and the distance to the market is greater than 44 km (Table 2). The lowest probability is predicted using the same variables at the opposite end of the thresholds. These results suggest that smallholders are likely to intensify their cropping systems by employ-

ing multi-season cropping when there is sufficient water availability, soil moisture, and access to highly commercialized markets. Additionally, labor scarcity and relatively higher annual per capita income decrease the likelihood of choosing multi-season cropping.

Regarding the choice of improved seed varieties, the analysis of this regression tree indicates that the probability of selecting improved seed varieties is highest when the household size is less than 5, the distance to the market is greater than 23 km, and the farm size is greater than 10 ha. Moreover, the lowest probability is predicted when a household has more than five members, its head is younger than 51 years old, it sells less than 72% of its crop yield, and it owns more than 9 ha of land. Interestingly, a larger farm size is a strong predictor in both situations. However, depending on how farm size interacts with other factors (e.g., household size and market access), it can

either increase or decrease the likelihood that improved genomes will be adopted.

Small-scale irrigation is most likely to be utilized if there are more than 110 person-days of available labor, farmers are not monoculture rice producers, and their farm has a TWI greater than 28. Similarly, having more than 110 labor days available, not being monoculture rice producers, having a TWI below 28.7, receiving less than 31% of income from sources other than farming, and having a household with more than five members all predict a lower probability. TWI (availability of sufficient water and soil moisture) strongly determines whether a farmer will choose small-scale irrigation, which is an intriguing aspect of this finding. The limited water availability, relatively large family size, and minimal participation in off-farm activities will decrease the likelihood of small-scale irrigation investment and use.

Additionally, there is a 27% chance that farm households with an area greater than 3.56 ha and a TWI less than 17.425 will use fertilizer. Those who cultivate less than 3.56 ha and live less than 35 km from the market will have the lowest probability (7%) of using fertilizer. The farm households with the highest probability (22%) to combine small-scale irrigation with fertilizer application have a nonfarm income greater than 38%, a farm size greater than 7 ha, and a market distance of less than 22 km.

Larger farms situated in drier portions of the floodplain are more likely to utilize fertilizer. Similarly, farmers who earn a significant portion of their income from non-farming sources, have large farm sizes, and reside further away from the market are more likely to apply fertilizer and irrigation on a smaller scale.

Finally, we looked qualitatively at how farmers in the valley perceived intensification limits. Farmers were asked if they planned to use any of the intensification options during the previous production season, and if not, they were prompted to explain the main obstacles or limitations. The main restrictions mentioned by farmers for each option are compared in Figure 7.

Most farmers cited the prohibitive price of fertilizer as the main deterrent to adoption (which account for 54.8% of non-users). Additionally, a belief that their soil is fertile (38.7%) and crop not suitable for fertilizers are additional reasons for non-fertilization. Regarding the use of improved seed, half of farmers cited the market's preference for conventional varieties as their main deterrent. Farmers gave additional justifications for not using improved seed varieties, including high transaction costs (9.47%), unaffordable prices (20.7%), a lack of financial capital (15.4%), and quality uncertainty (5.33%). The most common reason given for not using

irrigation are a lack of irrigation equipment (49.1%), restricted access to water (25.7%), enough water on the plot (21.6%), and a labor shortage (3.59%). Numerous factors, including inconsistent rainfall (51.1%) and a labor shortage, prevent the cultivation of multiple seasons (36.2 %).

### 6.3 | Limitations

Although we demonstrate the contributions of a data-driven Bayesian network approach to analyzing the underlying determinants of intensification choices in small-holder agriculture, this study has limitations, and our findings should be interpreted with these limitations in mind. As mentioned in the methodological approach section, some of these limitations are related to the general modeling approach, such as difficulty inferring a network structure from data, reliance on discrete variables, and high computational requirement.

There are also additional limitations that are peculiar to this study. First, although BBN is recommended for small and incomplete data sets (Uusitalo, 2007, p. 6; Kontkanen et al., 1997, p. 317) and our sample size is comparable to other studies (Prishchepov et al., 2019; Sun & Müller, 2013), it is small in comparison to the data used for typical machine learning problems. Second, the research was limited to a single floodplain, namely, KVF. Therefore, generalizations of the findings require first replicating these findings in other similar floodplain environments. Third, because other factors could not be accurately measured, those that may influence intensification decisions at the household level were excluded from this analysis (e.g., access to extension service). Fourth, rather than identifying causal relationships, we gain insights into how factors associated with intensification choices differ across options. Fifth, we cannot be certain that all relevant intensification options (other labor-saving intensification options, such as mechanization) were considered in the analysis because they were excluded from our survey instrument. Finally, the BBN presented in this study is static and does not account for dynamics. To this end, a planned follow-up study combines the BBN with a spatially explicit agent-based model (ABM) that considers farmer heterogeneity, interaction among themselves, and interaction with the floodplain. Combining BBN with an ABM reduces the computational challenges of ABMs by providing probabilistic agent rules. Moreover, representing BBN nodes as state variables in the ABM will provide temporal dynamics to the BBN approach (Kocabas & Dragicevic, 2013; Sun & Müller, 2013).



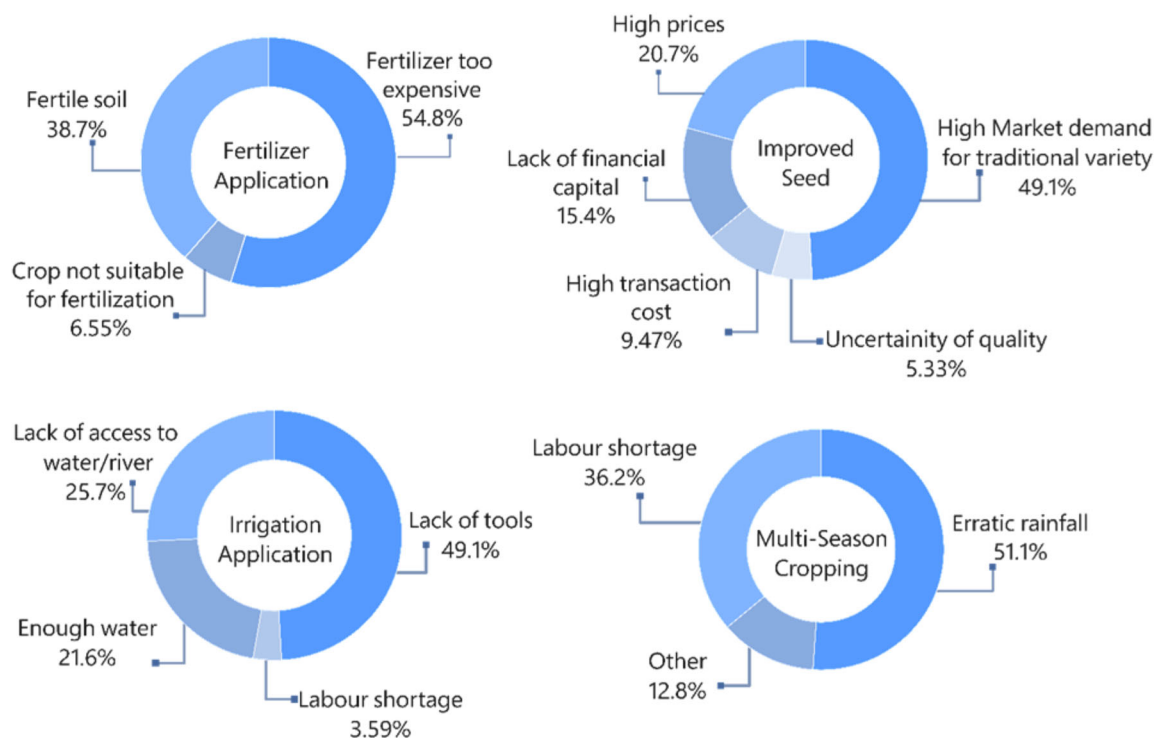


FIGURE 7 Perceived constraint for specific intensification option

## 7 | CONCLUSION

Increasing population pressure, efforts to protect the fragile wetland biome, and the need to increase food production will encourage smallholder farmers in the KVF to shift from land expansion to land-saving strategies, that is, intensification. Although intensification options have been gradually increasing (Otsuka & Larson, 2016), they are still not widely used, leaving significant room for further intensification by smallholders in the valley. This study sets out to systematically analyze the factors that influence the choice of various intensification options in the KVF's floodplain production systems. Our data-driven BBN allows us to represent complex interactions and dependencies between various factors and intensification choices while accounting for uncertainties based on statistical theory. Moreover, our augmented sensitivity analysis revealed individual choices for improved seed variety, fertilizer application, small-scale irrigation, and multi-season cropping. A variety of covariates were included, including plot and household characteristics, market access, and agro-ecological conditions. For better interpretability, we used a novel approach that combined a data-driven BBN with meta-modeling by regression trees. This revealed the relative importance of determinants in the selection of different strategies. The analyses show that the decision-making process for intensification options is complex and highly dependent on various factors. Although the

choice of each option is influenced differently by the covariates under consideration, access to nonfarm income, market access, and plot topography all play vital roles across options. Despite smallholder farmers' willingness to increase yield, those who did not use intensification options gave various reasons. The main reasons for not cropping multi-season are a lack of rain or water access and a labor shortage. Higher fertilizer prices and increased consumer demand for traditional seedlings are cited as the primary reasons for not using fertilizer, and improved seed—access to river networks and irrigation tools limits farmers' ability to irrigate their plots.

Our findings have important policy implications for targeted productivity enhancement in a vulnerable floodplain ecosystem and for explaining potential variation in intensification strategy choices. The preferences and choices of smallholders for various intensification pathways should be addressed in agricultural development programs and policies. These pathways must be tailored to site-specific conditions, local needs, and available resources to increase crop production in KVF. Fertilizer adoption can be increased if, for example, farmers with relatively large farm sizes and farms located on the peripheral (not in heavily flooded areas) are assisted in accessing fertilizer and encouraged to participate in the market simultaneously. Adoption of improved seed varieties would be boosted even further if nonfarm income streams and market access were made more accessible.

Similarly, increasing access to markets, equipment, and irrigation drainage systems for farmers in the floodplain's center may increase irrigation adoption.

Exploiting multiple cropping potential in KVF is a viable and underutilized option for increasing agricultural production and providing farmers with the opportunity to diversify, thus reducing risk or adding nutritional quality. Furthermore, such exploitation can mitigate the reduction of the regional water cycle and improve soil ecology (Xiang et al., 2021; Kawaski, 2019). This requires collective efforts that go far beyond simply promoting intensification through standardized extension services and increased input supply. It includes encouraging farmers to pursue the options that they believe are most appropriate in their specific circumstances.

Furthermore, incentivizing farmers to pursue intensification measures tailored to their specific needs and conditions could help reduce cropland encroachment in floodplains, which is currently the most serious problem.

## ACKNOWLEDGMENTS

The authors wish to thank Ashok Mishra, the associate editor and three anonymous reviewers who provided constructive feedback on the previous versions of the manuscript. B.G. and T.H. acknowledge support by the project "GlobeE: Wetlands in East Africa" funded by the German Federal Ministry of Education and Research (FKZ: 031A250 A-H) and the German Federal Ministry of Economic Cooperation and Development. B.G. and T.H. acknowledge additional funding by the German Research Foundation under the CRC/Transregio 228: Future Rural Africa: Future-making and social-ecological transformation (Project number: 328966760). T.H. and S.R. acknowledge funding by the German Research Foundation under Germany's Excellence Strategy, EXC-2070 - 390732324 - PhenoRob. All errors remain solely of the authors.

Open access funding enabled and organized by Projekt DEAL.

## REFERENCES

- Aalders, I. (2008). Modeling Land-Use Decision Behavior with Bayesian Belief Networks. *Ecology and Society*, 13(1), 1–16. <https://doi.org/10.5751/ES-02362-130116>
- Abay, K. A., Berhane, G., Taffesse, A. S., Koru, B., & Abay, K., & others. (2016). Understanding farmers' technology adoption decisions: Input complementarity and heterogeneity. *Ethiopia Strategy Support Program Working Paper*, 82(2), 1–22. <http://ebrary.ifpri.org/cdm/ref/collection/p15738coll2/id/130163>
- Agra (2016). *Africa Agriculture status report: Progress towards Agricultural Transformation in Africa*. Agra; Available at: <http://agra.org/aasr2016>
- Alomia-Hinojosa, V., Speelman, E. N., Thapa, A., Wei, H. - E., McDonald, A. J., Tittonell, P., & Groot, J. C. J. (2018). Exploring farmer perceptions of agricultural innovations for maize-legume intensification in the mid-hills region of Nepal. *International Journal of Agricultural Sustainability*, 16(1), 74–93. <https://doi.org/10.1080/14735903.2018.1423723>
- Amirkhani, H., Rahmati, M., Lucas, P. J. F., & Hommersom, A. (2017). Exploiting experts' knowledge for structure learning of Bayesian networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(11), 2154–2170. <https://doi.org/10.1109/TPAMI.2016.2636828>
- Bamford, A., Ferrol-Schulte, D., & Smith, H. (2010). *The status of the Ruipa corridor between the Selous Game Reserve and the Udzungwa Mountains*. Tanzania: Report to Frontier Tanzania, Dar Es Salaam.
- Besley, T., & Case, A. (1993). Modeling technology adoption in developing countries. *The American Economic Review*, 83(2), 396–402.
- Binswanger-Mkhize, H. P., & Savastano, S. (2017). Agricultural intensification: The status in six African countries. *Food Policy*, 67, 26–40. <https://doi.org/10.1016/j.foodpol.2016.09.021>
- Blattberg, R. C., Kim, B. - D., & Neslin, S. A. (2008). Machine learning. In R. C. Blattberg, B. - D. Kim, & S. A. Neslin (Eds.), *Database marketing: Analyzing and managing customers* (pp. 465–492). New York: Springer. [https://doi.org/10.1007/978-0-387-72579-6\\_19](https://doi.org/10.1007/978-0-387-72579-6_19)
- Cain, J. (2001). Improvements in natural resources water sector and beyond. *Ecology*, 44(0), 124.
- Celio, E., Koellner, T., & Grêt-Regamey, A. (2014). Modeling land use decisions with Bayesian networks: Spatially explicit analysis of driving forces on land use change. *Environmental Modelling and Software*, 52, 222–233. <https://doi.org/10.1016/j.envsoft.2013.10.014>
- Chen, S. H., & Pollino, C. A. (2012). Good practice in Bayesian Network Modelling. *Environmental Modelling and Software*, 37, 134–145. <https://doi.org/10.1016/j.envsoft.2012.03.012>
- Clarke, E. J., & Barton, B. A. (2000). Entropy and MDL discretization of continuous variables for Bayesian belief networks. *International Journal of Intelligent Systems*, 15(1), 61–92.
- Coutts, S. R., & Yokomizo, H. (2014). Meta-models as a straightforward approach to the sensitivity analysis of complex models. *Population Ecology*, 56(1), 7–19. <https://doi.org/10.1007/s10144-013-0422-1>
- Coulson, A. (2015). Small-scale and large-scale agriculture: Tanzanian experiences. In M. Stahl (Ed.), *Looking back, looking ahead—Land, agriculture and society in East Africa* (pp. 44–73). Nordic Africa Institute.
- Daly, R., Shen, Q., & Aitken, S. (2011). Learning Bayesian networks: Approaches and issues. *The Knowledge Engineering Review*, 26(2), 99–157.
- Dinesen, L. (2016). Kilombero Valley Floodplain (Tanzania). In C. M. Finlayson, G. R. Milton, R. C. Prentice, & N. C. Davidson (Eds.), *The Wetland Book: II: Distribution, Description and Conservation* (pp. 1–8). Springer.
- Dorfman, J. H. (1996). Modeling multiple adoption decisions in a joint framework. *American Journal of Agricultural Economics*, 78(3), 547–557.
- Drury, B., Valverde-Rebaza, J., Moura, M. - F., & de Andrade Lopes, A. (2017). A survey of the applications of Bayesian networks in agriculture. *Engineering Applications of Artificial Intelligence*, 65, 29–42.
- Erenstein, O. (2006). Intensification or extensification? Factors affecting technology use in peri-urban lowlands along an agro-ecological gradient in West Africa. *Agricultural Systems*, 90(1–3), 132–158. <https://doi.org/10.1016/j.agsy.2005.12.005>



- Feder, G., Just, R. E., & Zilberman, D. (1985). Adoption of agricultural innovations in developing countries: A survey. *Economic Development and Cultural Change*, 33(2), 255–298.
- Foster, A. D., & Rosenzweig, M. R. (2010). Microeconomics of technology adoption. *Annual Review of Economics*, 2(1), 395–424. <https://doi.org/10.1146/annurev.economics.102308.124433>
- Frank, S. K. (2015). Expert-based Bayesian Network modeling for environmental management [Univ., Inst. of Physical Geography]. In: *Frankfurt hydrology paper*. [Univ., Inst. of Physical Geography], 11. <http://publikationen.uni-frankfurt.de/frontdoor/index/index/docId/37645>
- Framer, J., Sun, Z., Müller, D., Munroe, D. K., & Xu, J. (2014). Analyzing the drivers of tree planting in Yunnan, China, with Bayesian networks. *Land Use Policy*, 36, 248–258.
- Friedman, N., Geiger, D., Goldszmidt, M., Division, C. S., Park, M., Langley, P., Smyth, P., Friedman, N., Geiger, D., & Goldszmidt, M. (1997). Bayesian network classifiers \*. *Machine Learning*, 163(2–3), 131–163.
- Gebrekidan, B. H., Heckelei, T., & Rasch, S. (2020). Characterizing farmers and farming system in Kilombero Valley Floodplain, Tanzania. *Sustainability*, 12(17), 7114. <https://doi.org/10.3390/su12177114>
- Genuer, R., Poggi, J. M., & Tuleau-Malot, C. (2010). Variable selection using random forests. *Pattern Recognition Letters*, 31(14), 2225–2236.
- Ghadim, A. K. A., Pannell, D. J., & Burton, M. P. (2005). Risk, uncertainty, and learning in adoption of a crop innovation. *Agricultural Economics*, 33(1), 1–9. <https://doi.org/10.1111/j.1574-0862.2005.00433.x>
- Gladwin, C. H. (1989). Modeling farmers' decisions to change: Using cognitive science in the design of agricultural technology. *Social Science Perspectives on Managing Agricultural Technology [Edited by Groenfeldt, D.]*, 86–22, 127–141.
- Gouvernement of Tanzania (2013). *2012 population and housing census, population distribution by administrative areas. Project report. National Bureau of Statistics. Government of Tanzania; Available at: http://repository.out.ac.tz/362/*
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning*. Springer New York. <https://doi.org/10.1007/978-0-387-84858-7>
- Headey, D., Dereje, M., & Taffesse, A. S. (2014). Land constraints and agricultural intensification in Ethiopia: A village-level analysis of high-potential areas. *Food Policy*, 48, 129–141. <https://doi.org/10.1016/j.foodpol.2014.01.008>
- Howley, P. O., Donoghue, C., & Heanue, K. (2012). Factors affecting farmers' adoption of agricultural innovations: A panel data analysis of the use of artificial insemination among dairy farmers in Ireland. *Journal of Agricultural Science*, 4(6), 171–179. <https://doi.org/10.5539/jas.v4n6p171>
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning (Vol., 103)*. Springer. <https://doi.org/10.1007/978-1-4614-7138-7>
- Jarvis, A., Reuter, H. I., Nelson, A., & Guevara, E. (2008). Hole-filled SRTM for the globe Version 4, available from the CGIAR-CSI SRTM 90m Database. <http://srtm.csi.cgiar.org>
- Jones, T., Bamford, A. J., Ferrol-Schulte, D., Hieronimo, P., McWilliam, N., & Rovero, F. (2012). Vanishing wildlife corridors and options for restoration: A case study from Tanzania. *Tropical Conservation Science*, 5(4), 463–474.
- Jun, C., Ban, Y., & Li, S. (2014). China: Open access to Earth land-cover map. *Nature*, 514(7523), 434.
- Kafle, B. (2010). Determinants of adoption of improved maize varieties in developing countries: A review. *International Research Journal of Applied and Basic Sciences*, 1(1), 1–7.
- Kajisa, K. (2016). On the determinants of low productivity of rice farming in Mozambique: Pathways to intensification. In: *In pursuit of an African Green Revolution* (pp. 13–38). Springer. [https://doi.org/10.1007/978-4-431-55693-0\\_2](https://doi.org/10.1007/978-4-431-55693-0_2)
- Kariyasa, K., & Dewi, Y. A. (2013). Analysis of factors affecting adoption of integrated crop management farmer field school (ICM-FFS) in swampy areas. *International Journal of Food and Agricultural Economics (IJFAEC)*, 1(1128-2016-92015), 29–38.
- Kassie, M., Marennya, P., Tessema, Y., Jaleta, M., Zeng, D., Erenstein, O., & Rahut, D. (2018). Measuring farm and market level economic impacts of improved maize production technologies in Ethiopia: Evidence from panel data. *Journal of Agricultural Economics*, 69(1), 76–95. <https://doi.org/10.1111/1477-9552.12221>
- Kassie, M., Teklewold, H., Jaleta, M., Marennya, P., & Erenstein, O. (2015). Understanding the adoption of a portfolio of sustainable intensification practices in eastern and southern Africa. *Land Use Policy*, 42, 400–411. <https://doi.org/10.1016/j.landusepol.2014.08.016>
- Kato, F. (2007). Development of a major rice cultivation area in March, 3–18.
- Kawasaki, K. (2019). Two harvests are better than one: Double cropping as a strategy for climate change adaptation. *American Journal of Agricultural Economics*, 101(1), 172–192.
- Khonje, M. G., Manda, J., Mkandawire, P., Tufa, A. H., & Alene, A. D. (2018). Adoption and welfare impacts of multiple agricultural technologies: Evidence from eastern Zambia. *Agricultural Economics*, 49(5), 599–609. <https://doi.org/10.1111/agec.12445>
- Kijima, Y. (2016). On the possibility of rice green revolution in rainfed areas in Uganda: Impact evaluation of a management training program and guidebook distribution. In *In Pursuit of an African Green Revolution* (pp. 65–89). Springer. [https://doi.org/10.1007/978-4-431-55693-0\\_4](https://doi.org/10.1007/978-4-431-55693-0_4)
- Kjaerulff, U. B., & Madsen, A. L. (2012). *Bayesian networks and influence diagrams: A guide to construction and analysis (Vol., 22)*. Springer Science & Business Media.
- Kocabas, V., & Dragicevic, S. (2013). Bayesian networks and agent-based modeling approach for urban land-use and population density change: A BNAS model. *Journal of Geographical Systems*, 15(4), 403–426. <https://doi.org/10.1007/s10109-012-0171-2>
- Koller, D., & Friedman, N. (2009). *Probabilistic graphical models: Principles and techniques*. MIT press.
- Kontkanen, P., Myllymaki, P., Silander, T., Tirri, H., & Grunwald, P. (1997). Comparing predictive inference methods for discrete domains. *Proceedings of the Sixth International Workshop on Artificial Intelligence and Statistics*, January, 311–318.
- Kopecký, M., Macek, M., & Wild, J. (2021). Topographic Wetness Index calculation guidelines based on measured soil moisture and plant species composition. *Science of the Total Environment*, 757, 143785. <https://doi.org/10.1016/j.scitotenv.2020.143785>
- Korb, K. B., & Nicholson, A. E. (2010). *Bayesian artificial intelligence*, (2nd ed.). CRC Press, Inc.

- Koski, T., & Noble, J. (2011). *Bayesian networks: An introduction* (Vol. 924). John Wiley & Sons.
- Kuhn, M., & Johnson, K. (2013). *Applied Predictive Modeling* (Vol. 810). Springer New York. <https://doi.org/10.1007/978-1-4614-6849-3>
- Kuhn, M. (2008). Caret package. *Journal of Statistical Software*, 28(5), 1–26.
- Kyrimi, E., McLachlan, S., Dube, K., Neves, M. R., Fahmi, A., & Fenton, N. (2021). A comprehensive scoping review of Bayesian networks in healthcare: Past, present and future. *Artificial Intelligence in Medicine*, 117, 102108. <https://doi.org/10.1016/j.artmed.2021.102108>
- Lee, D. (2005). Agricultural sustainability and technology adoption: Issues and policies for developing countries. *American Journal of Agricultural Economics*, 87(5), 1325–1334. <https://doi.org/10.1111/j.1467-8276.2005.00826.x>
- Leemhuis, C., Thonfeld, F., Näschen, K., Steinbach, S., Muro, J., Strauch, A., López, A., Daconto, G., Games, I., Diekkrüger, B., Leemhuis, C., Thonfeld, F., Näschen, K., Steinbach, S., Muro, J., Strauch, A., López, A., Daconto, G., Games, I., & Diekkrüger, B. (2017). Sustainability in the food-water-ecosystem nexus: The role of land use and land cover change for water resources and ecosystems in the Kilombero Wetland, Tanzania. *Sustainability*, 9(9), 1513. <https://doi.org/10.3390/su9091513>
- MALF (2015). Tanzania livelihood baseline profile: Kilombero-Ulanga-Lusewa paddy, maize, and cassava livelihood zone. MALF; (Issue 43),.
- Marcot, B. G., Steventon, J. D., Sutherland, G. D., & McCann, R. K. (2006). Guidelines for developing and updating Bayesian belief networks applied to ecological modeling and conservation. *Canadian Journal of Forest Research*, 36(12), 3063–3074. <https://doi.org/10.1139/X06-135>
- Margaritis, D. (2003). Learning Bayesian network model structure from data.
- McCartney, M., Rebelo, L. M., Senaratna Sellamuttu, S., & de Silva, S. (2010). Wetlands, agriculture and poverty reduction. <https://doi.org/10.5337/2010.230>
- Milder, J. C., Buck, L. E., & Hart, A. (2013). *Applying an agriculture green growth approach in the SAGCOT clusters: Challenges and opportunities in Kilombero*. Ithemi and Mbarali.
- Milder, J. C., Buck, L. E., Hart, A., Shames, S. A., Scherr, S. J., & Kozar, R. (2013). *Six opportunities to green agricultural production in the southern agricultural growth corridor of Tanzania (SAGCOT)*. Dar es Salaam: SAGCOT Centre.
- Milder, J. C., Buck, L. E., Hart, A. K., Scherr, S. J., & Shames, S. A. (2013). *A framework for agriculture green growth: Greenprint for the Southern agricultural growth corridor of Tanzania (SAGCOT)*. Dar Es Salaam: SAGCOT Centre.
- Molnar, C. (2019). Interpretable machine learning a guide for making black box models explainable. <http://leanpub.com/interpretable-machine-learning>
- Mombo, F., Speelman, S., Van Huylenbroeck, G., Hella, J., & Moe, S. (2011). Ratification of the Ramsar convention and sustainable wetlands management: Situation analysis of the Kilombero Valley wetlands in Tanzania. *Journal of Agricultural Extension and Rural Development*, 3(9), 153–164.
- Msofe, N. K., Sheng, L., Lyimo, J., Msofe, N. K., Sheng, L., & Lyimo, J. (2019). Land use change trends and their driving forces in the Kilombero Valley Floodplain, Southeastern Tanzania. *Sustainability*, 11(2), 505. <https://doi.org/10.3390/su11020505>
- Mwangi, M., & Kariuki, S. (2015). Factors determining adoption of new agricultural technology by smallholder farmers in developing countries. *Journal of Economics and Sustainable Development*, 6(5), 208–216.
- Mwimo, L., Mbowe, W., Kombe, C., Kibesse, B., Mziya, M., Kazi, M., Machemba, J., Mashini, S., Mpemba, Z., Lugobi, M., Yabu, N., & Ndunguru, E. (2016). Contract farming schemes in Tanzania: Benefits and challenges working paper series.
- Nagarajan, R., Scutari, M., & Lèbre, S. (2013). Bayesian networks in R. <https://doi.org/10.1007/978-1-4614-6446-4>
- Nakano, Y., Kajisa, K., & Otsuka, K. (2016). On the possibility of rice green revolution in irrigated and rainfed areas in Tanzania: An assessment of management training and credit programs. In *In pursuit of an African green revolution* (pp. 39–64). Japan: Springer. [https://doi.org/10.1007/978-4-431-55693-0\\_3](https://doi.org/10.1007/978-4-431-55693-0_3)
- Nakano, Y., Tanaka, Y., & Otsuka, K. (2018). Impact of training on the intensification of rice farming: Evidence from rainfed areas in Tanzania. *Agricultural Economics*, 49(2), 193–202. <https://doi.org/10.1111/agec.12408>
- Neapolitan, R. E. (2010). Learning Bayesian networks. *International Journal of Data Mining and Bioinformatics*, 38, 505–519. <https://doi.org/10.1016/j.jbi.2010.03.005>
- Nielsen, T. D., & Jensen, F. V. (2009). *Bayesian networks and decision graphs*. Springer Science & Business Media.
- Nindi, S. J., Maliti, H., Bakari, S., Kija, H., & Machoke, M. (2014). Conflicts over land and water resources in the Kilombero Valley floodplain, Tanzania. *African Study Monographs*, 50(October), 173–190.
- Nkonya, E., Schroeder, T., & Norman, D. (1997). Factors affecting adoption of improved maize seed and fertiliser in northern Tanzania. *Journal of Agricultural Economics*, 48(1–3), 1–12. <https://doi.org/10.1111/j.1477-9552.1997.tb01126.x>
- Nojavan, A. F., Qian, S. S., & Stow, C. A. (2017). Comparative analysis of discretization methods in Bayesian networks. *Environmental Modelling and Software*, 87, 64–71. <https://doi.org/10.1016/j.envsoft.2016.10.007>
- NorsysSoftwareCorp (2016). *Netica-J manual*. NorsysSoftwareCorp.
- Ntongani, W. A., Munishi, P. K. T., More, S. R., & Kashaigili, J. J. (2014). Local knowledge on the influence of land use/cover changes and conservation threats on avian community in the Kilombero Wetlands, Tanzania. *Open Journal of Ecology*, 4(September), 723.
- Ntshangase, N. L., Muroyiwa, B., & Sibanda, M. (2018). Farmers' perceptions and factors influencing the adoption of no-till conservation agriculture by small-scale farmers in Zashuke, KwaZulu-Natal province. *Sustainability (Switzerland)*, 10(2), 555. <https://doi.org/10.3390/su10020555>
- Okike, I., Jabbar, M. A., Manyong, V., Smith, J. W., Akinwumi, J. A., & Ehui, S. K. (2001). Agricultural intensification and efficiency in the West African savannahs: Evidence from northern Nigeria. *Socio-Economics and Policy Research Working Paper - International Livestock Research Institute*, 33, 1–54. <https://doi.org/10.22004/ag.econ.182891>
- Otsuka, K., & Larson, D. F. (Eds.). (2016). *In pursuit of an African Green Revolution*. Springer. <https://doi.org/10.1007/978-4-431-55693-0>

- Otsuka, K., & Place, F. (2013). *Evolutionary changes in land tenure and agricultural intensification in sub-Saharan Africa*. United Nations University Working Paper, 2014/051(November), 47. <https://doi.org/10.1093/oxfordhb/9780199687107.013.015>
- Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. Morgan Kaufmann.
- Pearl, J. (2009). Causality: Models, reasoning and inference. In: *International statistical review* (2nd ed., Vol., 79, (Issue 2), Cambridge University Press.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., & Dubourg, V., & others. (2011). Scikit-learn: Machine learning in Python. *Of Machine Learning*, 12(Oct), 2825–2830. <https://doi.org/10.1007/s13398-014-0173-7.2>
- Perz, S. G. (2003). Social determinants and land use correlates of agricultural technology adoption in a forest frontier: A Case Study in the Brazilian Amazon, 31(1), 133–165. <http://doi.org/10.1023/A:1022838325166>
- Pollino, C. A., Woodberry, O., Nicholson, A., Korb, K., & Hart, B. T. (2007). Parameterisation and evaluation of a Bayesian network for use in an ecological risk assessment. *Environmental Modelling and Software*, 22(8), 1140–1152. <https://doi.org/10.1016/j.envsoft.2006.03.006>
- Pope, A., & Gimblett, R. (2017). Linking participatory, Bayesian, and agent-based modeling techniques to simulate coupled natural-human system: A Case Study with Ranchers in Sonora, Mexico. In *Environmental modeling with stakeholders* (pp. 345–359). Springer.
- Prishchepov, A. V., Ponkina, E., Sun, Z., & Müller, D. (2019). Revealing the determinants of wheat yields in the Siberian breadbasket of Russia with Bayesian networks. *Land Use Policy*, 80, 21–31. <https://doi.org/10.1016/j.landusepol.2018.09.038>
- Pronti, A., Auci, S., & Mazzanti, M. (2020). *What are the factors driving the adoption and intensity of sustainable irrigation technologies in Italy*. CERCIS Work. Pap. Ser, 1–44.
- Ragno, E., Hrachowitz, M., & Morales-Nápoles, O. (2022). Applying non-parametric Bayesian networks to estimate maximum daily river discharge: potential and challenges. *Hydrology and Earth System Sciences*, 26(6), 1695–1711. <https://doi.org/10.5194/hess-26-1695-2022>
- Rasmussen, S., Madsen, A. L., & Lund, M. (2013). Bayesian network as a modelling tool for risk management in agriculture, 1–14.
- Reardon, T., Stamoulis, K., & Pingali, P. (2007). Rural nonfarm employment in developing countries in an era of globalization. *Agricultural Economics*, 37, 173–183.
- Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why Should I Trust You?": Explaining the Predictions of Any Classifier. <http://arxiv.org/abs/1602.04938>
- Rogers, J., & Gunn, S. (2006). *Identifying feature relevance using a random forest* (pp. 173–184). Springer. [https://doi.org/10.1007/11752790\\_12](https://doi.org/10.1007/11752790_12)
- Rogers, J. D., & Gunn, S. R. (2005). Ensemble algorithms for feature selection (pp. 180–198). [https://doi.org/10.1007/11559887\\_11](https://doi.org/10.1007/11559887_11)
- Rohmer, J., & Gehl, P. (2020). Sensitivity analysis of Bayesian networks to parameters of the conditional probability model using a Beta regression approach. *Expert Systems with Applications*, 145, 113130. <https://doi.org/10.1016/j.eswa.2019.113130>
- Sanchez, S. M. (2005). NOLHdesigns spreadsheet. Software Available Online via <http://Diana.Cs.Nps.Navy.Mil/SeedLab>, 1430–1439.
- Salasya, B., Mwangi, W. M., Mwabu, D., & Diallo, A. (2007). Factors influencing adoption of stress-tolerant maize hybrid (WH 502) in western Kenya.
- Sandri, M., & Zuccolotto, P. (2006). Variable selection using random forests. In S. Zani, A. Cerioli, M. Riani, & M. Vichi (Eds.), *Data analysis, classification and the forward search* (pp. 263–270). Springer. [https://doi.org/10.1007/3-540-35978-8\\_30](https://doi.org/10.1007/3-540-35978-8_30)
- Schelhas, J. (1996). Land use choice and change: Intensification and diversification in the lowland tropics of Costa Rica. *Human Organization*, 55(3), 298–306. <https://doi.org/10.17730/humo.55.3.k4v1050710073063>
- Senthilkumar, K., Tesha, B. J., Mghase, J., & Rodenburg, J. (2018). Increasing paddy yields and improving farm management: Results from participatory experiments with good agricultural practices (GAP) in Tanzania. *Paddy and Water Environment*, 16(4), 749–766. <https://doi.org/10.1007/s10333-018-0666-7>
- Sheahan, M., & Barrett, C. B. (2014). Understanding the Agricultural Input Landscape in Sub-Saharan Africa: Recent Plot, Household, and Community-Level Evidence August. <https://doi.org/10.1596/1813-9450-7014>
- Shriar, A. J. (2000). Agricultural intensity and its measurement in frontier regions. *Agroforestry Systems*, 49(3), 301–318. <https://doi.org/10.1023/A:1006316131781>
- Shriar, A. J. (2001). The dynamics of agricultural intensification and resource conservation in the buffer zone of the Maya Biosphere Reserve, Petén, Guatemala. *Human Ecology*, 29(1), 27–48.
- Sirén, A. H. (2007). Population growth and land use intensification in a subsistence-based indigenous community in the Amazon. *Human Ecology*, 35(6), 669–680. <https://doi.org/10.1007/s10745-006-9089-y>
- Sörensen, R., Zinko, U., & Seibert, J. (2006). On the calculation of the topographic wetness index: Evaluation of different methods based on field observations. *Hydrology and Earth System Sciences Discussions*, 10(1), 101–112.
- Sun, Z., & Müller, D. (2013). A framework for modeling payments for ecosystem services with agent-based models, Bayesian belief networks and opinion dynamics models. *Environmental Modelling and Software*, 45, 15–28. <https://doi.org/10.1016/j.envsoft.2012.06.007>
- Teklewold, H., Kassie, M., Shiferaw, B., & Köhlin, G. (2013). Cropping system diversification, conservation tillage and modern seed adoption in Ethiopia: Impacts on household income, agrochemical use and demand for labor. *Ecological Economics*, 93, 85–93. <https://doi.org/10.1016/j.ecolecon.2013.05.002>
- Ticehurst, J. L., Curtis, A., & Merritt, W. S. (2011). Using Bayesian Networks to complement conventional analyses to explore landholder management of native vegetation. *Environmental Modelling and Software*, 26(1), 52–65. <https://doi.org/10.1016/j.envsoft.2010.03.032>
- URT (2013). *National agriculture policy*. URT; Available at: <https://www.kilimo.go.tz/index.php/en/resources/view/national-agriculture-policy-2013>
- USAID (2019). *Agriculture factsheet*: Tanzania: USAID; Available at: [https://www.usaid.gov/sites/default/files/documents/1860/Agriculture%7B%5C\\_%7DFact%7B%5C\\_%7DSheet%7B%5C\\_%7DApril%7B%5C\\_%7D2019%7B%5C\\_%7DFINAL.pdf](https://www.usaid.gov/sites/default/files/documents/1860/Agriculture%7B%5C_%7DFact%7B%5C_%7DSheet%7B%5C_%7DApril%7B%5C_%7D2019%7B%5C_%7DFINAL.pdf)
- Uusitalo, L. (2007). Advantages and challenges of Bayesian networks in environmental modelling. *Ecological Modelling*, 203(3-4), 312–318. <https://doi.org/10.1016/j.ecolmodel.2006.11.033>

- Vanlauwe, B. (2016). *Challenges and opportunities for agricultural intensification of the humid highland systems of Sub-Saharan Africa*. Springer International Pu.
- Vanlauwe, B., Coyne, D., Gockowski, J., Hauser, S., Huising, J., Masso, C., Nziguheba, G., Schut, M., & Van Asten, P. (2014). Sustainable intensification and the African smallholder farmer. *Current Opinion in Environmental Sustainability*, 8, 15–22. <https://doi.org/10.1016/j.cosust.2014.06.001>
- Wainaina, P., Tongruksawattana, S., & Qaim, M. (2016). Trade-offs and complementarities in the adoption of improved seeds, fertilizer, and natural resource management technologies in Kenya. *Agricultural Economics*, 47(3), 351–362. <https://doi.org/10.1111/agec.12235>
- WFP (2019). *WFP Tanzania country brief*. WFP; Available at: [https://docs.wfp.org/api/documents/WFP-0000105875/download/?\\_ga=2.160559294.2017491598.1566509334-1454385796.1566509334](https://docs.wfp.org/api/documents/WFP-0000105875/download/?_ga=2.160559294.2017491598.1566509334-1454385796.1566509334)
- Xiang, M., Yu, Q., Li, Y., Shi, Z., & Wu, W. (2022). Increasing multiple cropping for land use intensification: The role of crop choice. *Land Use Policy*, 112, 105846. <https://doi.org/10.1016/j.landusepol.2021.105846>
- Xie, H., & Huang, Y. (2021). Influencing factors of farmers' adoption of pro-environmental agricultural technologies in China: Meta-analysis. *Land Use Policy*, 109, 105622.
- Yamano, T., Rajendran, S., & Malabayabas, M. L. (2015). Farmers' self-perception toward agricultural technology adoption: Evidence on adoption of submergence-tolerant rice in Eastern India. *Journal of Social and Economic Development*, 17(2), 260–274. <https://doi.org/10.1007/s40847-015-0008-1>
- Yesuf, M., & Köhlin, G. (2008). *Market imperfections and farm technology adoption decisions: A case study from the highlands of Ethiopia*. Gothenburg, Sweden: environment for Development Discussion Paper Series 4.
- Zheng, F., & Webb, G. I. (2010). Tree augmented naive Bayes. In: C. Sammut & G. I. Webb (Eds.), *Encyclopedia of machine learning* (pp. 990–991). US: Springer.

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Gebrekidan, B. H., Heckelei, T., & Rasch, S. (2023). Modeling intensification decisions in the Kilombero Valley floodplain: A Bayesian belief network approach. *Agricultural Economics*, 54, 23–43. <https://doi.org/10.1111/agec.12740>