Ecological Informatics 79 (2024) 102434

Contents lists available at ScienceDirect

Ecological Informatics

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

Inferring multiple coffee flowerings in Central America using farmer data in a probabilistic model

Leonel Lara-Estrada [a,*], Luis Enrique Sucar [b], Livia Rasche [c,d]

[a] Natural Resources Institute, University of Greenwich, Chatham ME4 4TB, UK
[b] Instituto Nacional de Astrofísica, Óptica y Electrónica, Puebla, Mexico
[c] Research Unit Sustainability and Climate Risks, Universität Hamburg, Germany
[d] Chair of Land Use Economics, University of Hohenheim, Germany

ARTICLE INFO

Keywords:
Coffee arabica
Farmers' data
Ecological modeling
Climate risk
Dynamic Bayesian network

ABSTRACT

Coffee (*Coffea arabica* L.) is a climate-sensitive crop; rainfalls may trigger flowering event occurrences, and extreme rainfall during a flowering day can cause considerable yield reductions. Multiple flowering events can occur in the span of 12 months; the number varies from year to year. This paper introduces a Bayesian network model capable of inferring coffee flowering events in coffee areas in the Pacific Region of Central America based on observed data for coffee flowering and precipitation. The model structure was determined based on expert knowledge, and the model parametrization was learned from 53 years of data registered in the region. Data from four farms in the region were used for model validation. The model's performance in the inference of flowering intensity was good (spherical payoff of 0.78 out of maximal 1.00), and the model was able to depict expected behaviors for single and multiple flowerings. Further, comprehensive new details on the dynamics of multiple flowerings within a crop season were obtained, e.g., that a large flowering event tends to occur more quickly (8 to 10 days) after rain than a small flowering (10 to 13 days). We believe that this Bayesian network model has the potential to evolve and support the development of agricultural index-based insurance to deal with yield losses due to extreme rainfall during flowering. The use of longer farm records for model building may also serve to increase farmers' trust in the reliability of the tool.

1. Introduction

With a global export worth US$ 19 billion in 2017, coffee is one of the most important agricultural commodities (Bozzola et al., 2021). Producers are mostly smallholder farmers who grow the crop on over 12 million farms across 20 main countries (e.g., Ethiopia, Indonesia, Colombia, and others); in many of these countries, coffee is of high economic importance and can account for up to 20% of national export revenues (Bozzola et al., 2021). While the demand for coffee is rising steadily and the global coffee sector is expanding, the climatic conditions for coffee production are becoming less suitable in many areas (Lara-Estrada et al., 2021). This is particularly the case for the predominant coffee species *Coffea arabica* L., which has a narrow climatic niche in which high coffee quality and yields can be produced. Temperatures should be between 18 and 22 °C year-round, with night temperatures not lower than 15 °C, and annual precipitation sum should be between 1500 and 2000 mm, with at least one dry period of three to four months and preferably an even rainfall distribution in the wet season (Bertrand et al., 2012; Descroix and Snoeck, 2004; Lara-Estrada et al., 2017). Higher temperatures and untimely or insufficient precipitation can influence the coffee plants in the vegetative and reproductive phases and negatively affect flowering, fruiting, and bean quality, and thus yields and profits (Gay et al., 2006; Lin, 2007; Muschler, 2001).

Climate projections for Central America indicate that temperatures will rise, precipitation patterns will change, and extreme weather events will increase in the coming decades. This may lead to a reduction in the land suitability for coffee production and consequently to a decline in the quantity and quality of ecosystem services provided by coffee areas (e.g., pollinators, carbon sequestration) (Imbach et al., 2018, 2017; Lara-Estrada et al., 2023, 2021; Liu et al., 2023). Elevated temperatures can lead to premature ripening of coffee berries, thus decreasing their quality (Bertrand et al., 2012). Continuous heavy rainfall during the harvesting period can also affect the coffee quality and yields by increasing the risk of coffee bean defects, bean dropping, or a crack in

* Corresponding author.
E-mail address: L.LaraEstrada@greenwich.ac.uk (L. Lara-Estrada).

https://doi.org/10.1016/j.ecoinf.2023.102434
Received 31 August 2023; Received in revised form 7 November 2023; Accepted 12 December 2023
Available online 15 December 2023
1574-9541/© 2023 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

the bean skin (Kath et al., 2021; Murugan et al., 2022).

A coffee flowering occurs after a rainfall following a dry period, as the buds lie dormant while the plant experiences water stress. Rainfall events break this water stress, reactivating the development of buds, and the opening of the flowers (anthesis) some days later (Fig. 1) (Alvim, 1960). If, in these following days, the precipitation is minimal or absent, the water stress increases until the next rainfall event, where a new flowering event may be triggered. The intensity of the flowering (number of flowers) will depend on the severity of the accumulated water stress the coffee plant experiences and the amount of rainfall that breaks the stress (Alvim, 1960; Pagotto Ronchi and Rodrigues Miranda, 2020; Schuch et al., 1992). Irrigation can be used to produce larger flowering events and uniform coffee berry ripening and, therefore, a shorter harvesting period (Goodyear, 2004; Masarirambi et al., 2009).

If rains occur on the actual flowering day, some flowers may be damaged and not develop into fruits. This will impact yield, which will be decreased to a lesser or larger extent, depending on the intensity of flowering and rainfall (Lara-Estrada et al., 2012; Murugan et al., 2022). Therefore, flowering is one of the most anticipated phenological stages for coffee farmers because the number of flowering events and intensity are indicators of the harvest duration and final yields. Any disturbance during flowering may lead to yield losses, so farmers avoid having farm workers on the plantation during this time [Author personal observation].

Most coffee producers in Central America are smallholder farmers who depend on a good coffee harvest for their livelihoods (Bacon, 2005; Osorio, 2002). These farmers can neither influence the weather patterns nor adapt to climatically more extreme years by changing their crops for the season, as coffee plants are perennials and only renovated after decades. One option for farmers to decrease the risks of high financial losses is the use of agricultural index-based insurance (Clement et al., 2018; Eze et al., 2020). The advantages of such insurances are that they are attractive for all farmers, not only those who are more likely to suffer yield losses, that they are easy to implement, and that insurance fraud is difficult due to the use of readily available, verifiable, and accepted indices. Regarding the risk associated with rainfall during flowering, the first step would be to create a model to predict the flowering events and their intensity. In existing coffee crop models, coffee flowering is either modeled as a single event in the year after a rainfall threshold is reached (van Oijen et al., 2010), or as two flowering events using degree days accumulation between phenological stages as trigger parameters (Montoya-Restrepo et al., 2009). These simplified depictions of coffee flowering events do not reflect the more varied dynamics observed on coffee farms, with fluctuating numbers and intensities of flowering events from year to year.

This paper, therefore, introduces a probabilistic coffee flowering model to infer multiple flowering events based on the month and the rainfall data. The model is a Bayesian Network (BN) based on observed data on the flowering events and rainfall data recorded over decades by farmers in the Pacific Region of Nicaragua. The resulting BN model was validated and used to produce new insights into the flowering dynamics. Its graphical nature creates a sophisticated but still user-friendly tool with the potential to evolve into a model that can be used in a climate index insurance scheme to support smallholder farmers in coffee-producing countries by strengthening their climate resilience. Such a flowering model could also help farmers to synchronize flowerings and harvesting periods using irrigation (Masarirambi et al., 2009) or identify the optimal flowering period (Liu et al., 2020). To our knowledge, this is the first model of this kind.

2. Materials and methods

2.1. Study region and data

For the building and testing of the flowering model, we used previously collected data from *Coffeea arabica* plantations in the Pacific region coffee areas of Nicaragua (Fig. 2) (Lara-Estrada et al., 2012). These data include the dates and the intensity of flowerings, as well as records of daily precipitation (Fig. 3). The study region's biophysical conditions are representative of Pacific areas over Central American countries, experiencing a well-defined dry season with the lowest rate of rainfall and high temperatures in the region (Bornemisza et al., 1999; Hidalgo et al., 2017; Taylor and Alfaro, 2005).

The data used for model training and validation was collected from four coffee farms: *San Francisco* and *El Rosal*, both located near the city of San Marcos in the Province Carazo, and *San Jose* and *Jardin Botanico*,

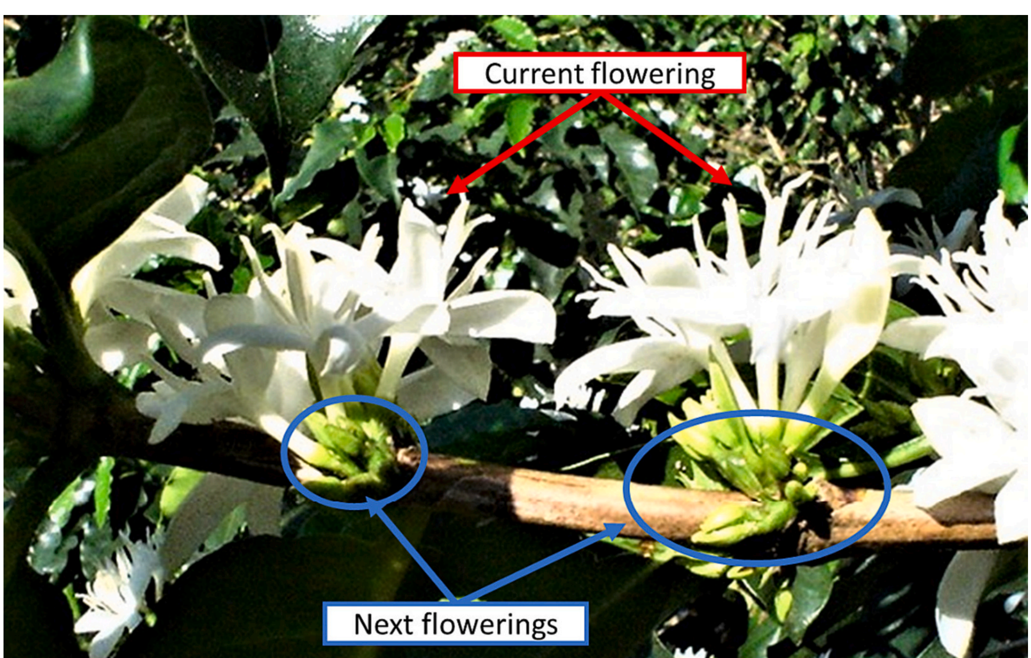

Fig. 1. Buds and flowers of the coffee plant. While some buds have already reached anthesis, not all buds break their dormancy after a rainfall. Smaller buds for the next flowering event are already laid.

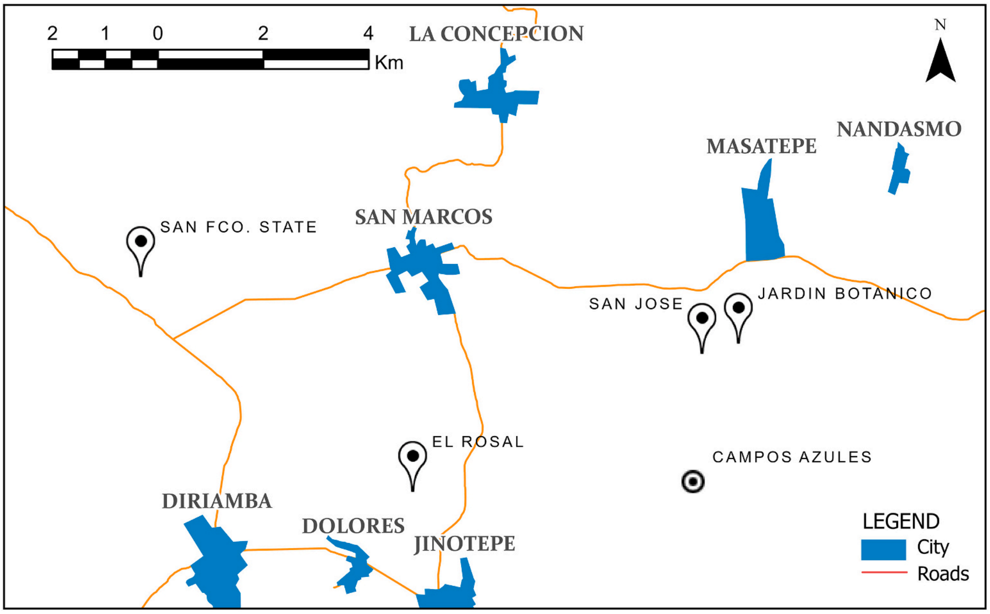

Fig. 2. Location of the coffee farms surveyed in the study region in the Pacific Region of Nicaragua.

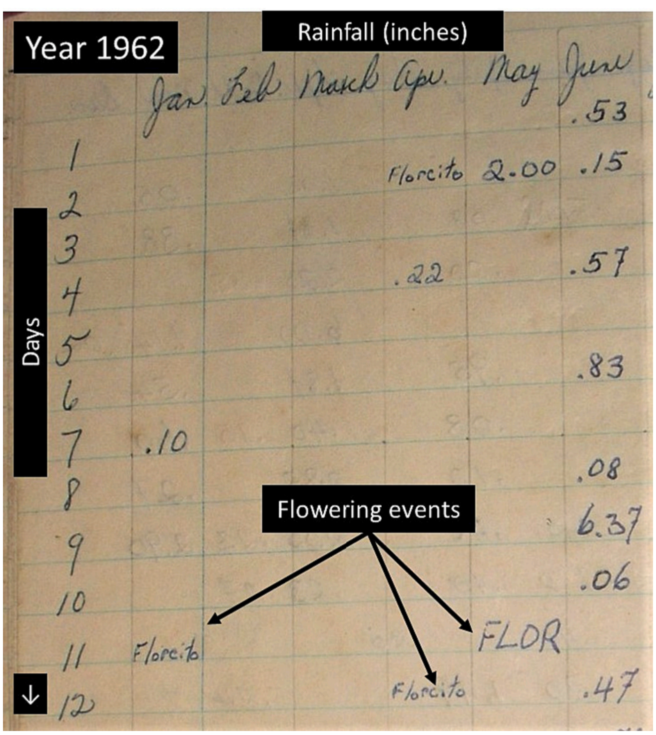

Fig. 3. Example of the rainfall and coffee flowering data registers used for model building and validation. *FLOR* indicates a large flowering event, and *Florcita* a small one.

both located near the city of Masatepe in the Province of Masaya, Nicaragua. We split the data to obtain two independent datasets, one for model training and one for model validation. For model training, we used 53 years of data from the farm *San Francisco* corresponding to 1943–1998 (1980 and 1981 were missing) (Fig. 3). For the model validation, we used four years of data (1999–2003) from *San Francisco*, as well as the records of the following years for each of the other three farms: *San Jose*: 2001–2002, 2004, 2006, 2008–2010; *El Rosal*: 2005–2010; *Jardin Botanico*: 2003, 2005–2007, 2009–2010. The years were chosen based on flowering data availability. Unfortunately, only the farm *San Francisco* had complete records of daily rainfall, which is why data from this farm was chosen for the model training. Rainfall data from the weather station at the Research Station *Campos Azules* was used for the three other farms; the research station is located between 2 and 4 km from the farms (Fig. 2).

The datasets were used to create the variables, *flowering intensity*, *month*, *rainfall inducing flowering* (amount of rainfall that induces a flowering), and *days to flowering after rain*. The variable *month* depicts the month where flowering occurs and acts as a proxy variable for the accumulated water stress observed in the region. A water stress index would have required more input data and variables for calculation and been more complicated to estimate and use for practitioners (Kögler and Söffker, 2017). Fig. 4 displays the monthly rainfall and air temperature observed in the region study region. The soil water stock experiences a gradual depletion during the dry period from January to April; depending on the accumulated water stress, if rainfall occurs during that time, the crop water stress would be broken, and a flowering event triggered (Goodyear, 2004).

2.2. Building and training the Bayesian network model

Bayesian Networks (BN) are multivariate statistical models that consist of two main components: a directed acyclic graph (model structure) and local conditional probability distributions (model parameters); together, they compactly represent the joint probability distribution. The acyclic graph consists of a set of (nodes) variables linked by arcs, the direction of which defines the conditional dependencies (parent node ➔ child node). However, once two variables are linked, even if the arrow goes in one direction, the inference flows in both directions, forward and backward, so by knowing the parent's state, the child's state can be inferred, and vice versa (Sucar, 2015a; Uusitalo, 2007). Variables can be discrete or continuous; in the case of continuous variables, they could be represented using parametric distributions, in particular Gaussian, or could be discretized in at least two mutually exclusive states. The structure of the directed acyclic graph can either be determined by the user (expert knowledge), learned from data, or a combination. The conditional probability tables (CPT) quantify the dependencies between the variables, which means that given the state of the parent node(s), the occurrence of a specific state in the child node has a certain probability. Therefore, for each variable in the BN, a CPT

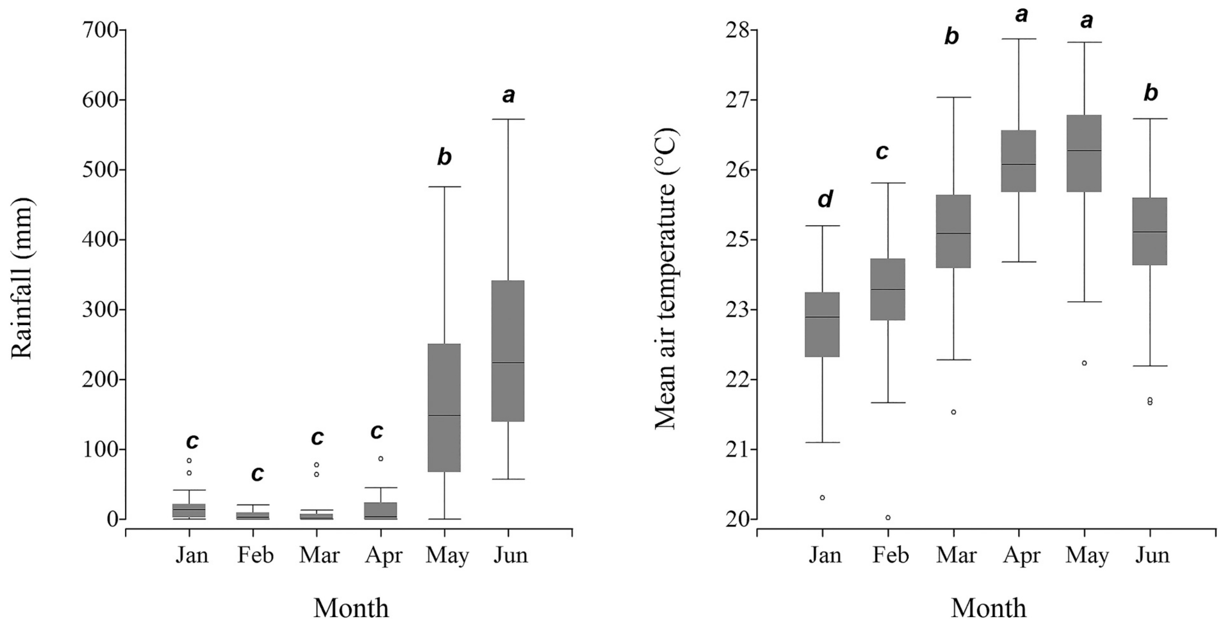

Fig. 4. Mean values for rainfall (1943–1998) and air temperature (1984, 1987, 1990, 1997, 1999) reported for the study region. Rainfall data comes from the farm *San Francisco*; and air temperature from the Research Station *Campos Azules*. Means with a common letter are not significantly different ($p > 0.05$).

has to be specified for the node variable given its parents in the graph; for root nodes (no parents), their marginal prior probability is specified (Aguilera et al., 2011; Uusitalo, 2007). These features have made BNs a suitable modeling option to depict, understand, and predict environmental and ecological relationships and processes in (agro)ecosystems (Aguilera et al., 2011; Beall et al., 2022; Hui et al., 2022).

The primary purpose of this model is to evaluate: If a rainfall event occurred after a period of water stress, 1) how large will the flowering event be? and 2) when will it occur? Based on expert knowledge and scientific literature, we defined the graphical model structure using the software Netica v.6.09 (Norsys Software Corp.). The datasets were used to create the variables and then added into the graphical model structure as node variables: *flowering intensity*, *days to flowering after rain, rainfall inducing flowering* (amount of rainfall that induces a flowering), and *month.* The variable *flowering intensity* was connected to the nodes *month* and *rainfall inducing flowering,* which were connected to each other and to the node *days to flowering* (Fig. 5, top four boxes). This basic pattern of four interconnected variables represents a single flowering (F) event that was repeated three more times to depict four possible flowering events in a given crop year (in >90% of the years, one to four flowering events occurred). The link from one pattern to the next was created through *flowering intensity*, as the number of flowering events in a year is an additional factor that influences flowering intensities (e.g., if there is only one flowering event in a year, it is a large one). If all four flowering events (F1-F4) take place, they occur at different time slices where each event gets feedback from the others to depict a Dynamic Bayesian Network (Sucar, 2015b; Uusitalo et al., 2018). Considering the forward-backward inference properties of BNs, in the model, the node *flowering intensity* is the parent of *month* and *rainfall inducing flowering.* This does not denote an actual physical causal relationship, only a statistical dependency. We chose to structure the model in this way to keep the CPTs simple (minimum number of incoming connections per node) and allow for a sole connection of one flowering event to the next through the variable *flowering intensity* (Marcot, 2017).

Once the structure of the model was determined (nodes linked), the state values for each variable were defined. We discretized the values of all variables into the following states: Large, Small, and No (flowering) for *flowering intensity*; January to June for *Month,* which are the months when flowerings were reported in the training dataset; 0–2.5 mm, 2.5–5 mm, 5–10 mm, ≥10 mm for *rainfall inducing flowering*; and 1–8 days, 8–10 days, 10–13 days for *days to flowering after rain.* The variables' states were defined based on statistical analysis for continuous variables and the existing possible values available for the discrete ones. After all variables' states were defined, we ran the Counting-Learning Algorithm (Norsys, 2023) using the training dataset to populate the CPTs of each node. The algorithm uses the training data to modify the conditional probabilities of the nodes, which all start with uniform CPTs. Every new case – e.g., a flowering event on a specific date with a specific size (small, large) after a specific amount of rainfall – updates only the CPTs of the nodes for which the case provides values. In case of missing combinations of variables' states in the dataset, the model will produce a uniform distribution, which will be used during the Bayesian inference if that case is requested. Once the model is trained and compiled, it is ready to use. Fig. 5 shows the compiled model without evidence entered; the numbers depict the variables' values according to the training data, displaying the overall observed dynamic of flowering in the region for the first time. It can be observed, for example, that three flowering events are most frequent, that it is most probable that the first flowering to be small (53.7% probability) and in February (35.2%), the second and third flowerings to be large and occur in May.

2.3. Model assumptions

Only four flowering events per year are modeled (90% of the observed cases), but more may occur with lower frequency. There is a finite number of possible flowers per year, which is triggered by the alternation of an accumulated water stress period and sufficient rainfall to break it. Once the plant water stress ends with the establishment of the rainy season, no more flowering events will occur. If there is a single flowering, it will be large. Possible changes in the climate patterns for 1944–1998 and their effects on the flowering dynamics of the study region are captured in the data and, therefore, in the model's priors and posterior inferences. Because there is no information available on the coffee varieties used by farmers, it is assumed that they used the same or that varietal differences have no effect on flowering. The model could be extended to include >4 flowering events if needed. The coffee varieties reported during the data collection were Caturra, Catuai, Pacas, and Bourbon (Lara-Estrada et al., 2012); we assumed no differences between them in their response to the triggering flowering factors due to their high genetic proximity (Anthony et al., 2002; Montagnon et al., 2012).

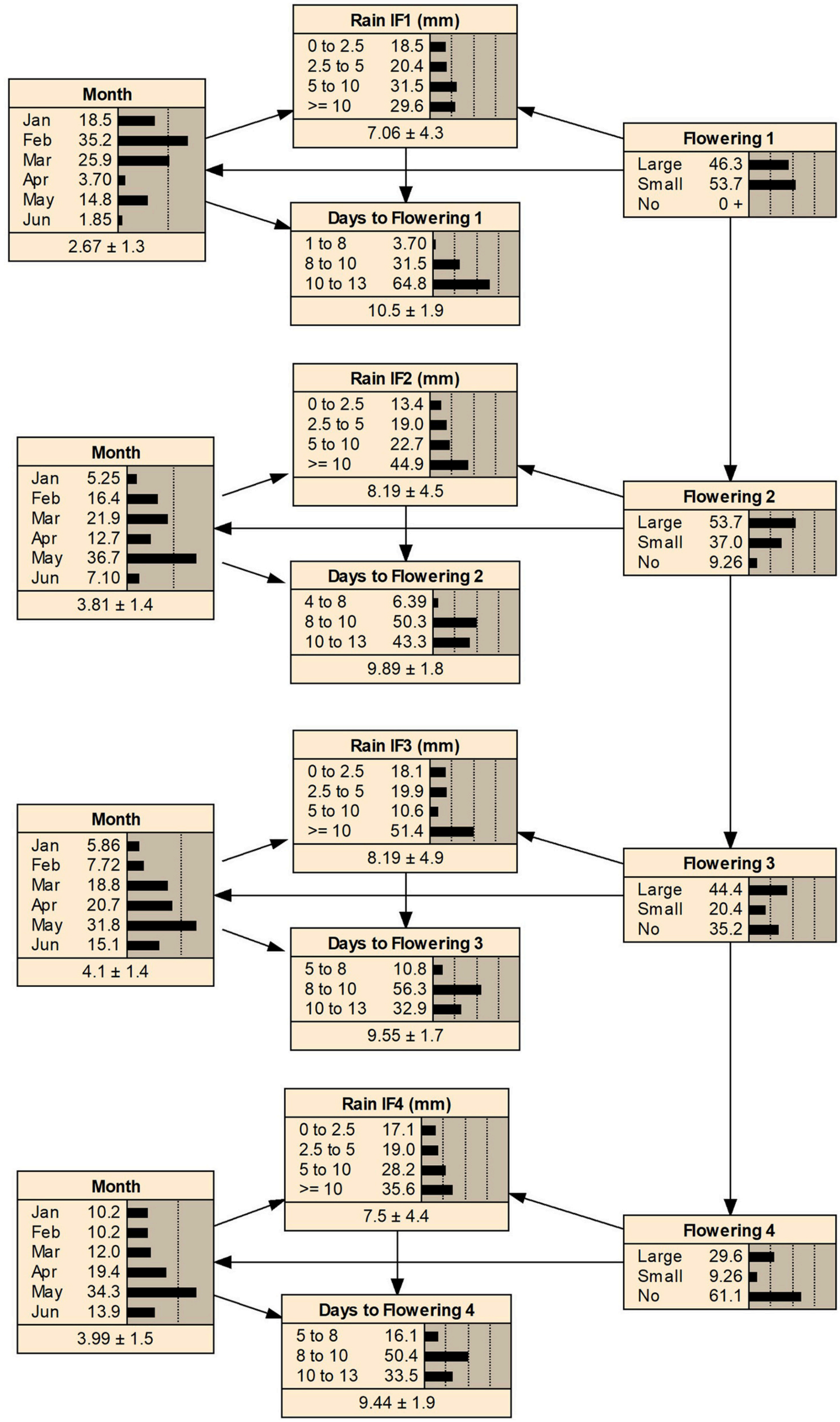

Fig. 5. Coffee flowering model. A maximum of four flowering events per year are possible. The model depicts the prior probabilities for each variable prior to data entery. If a user enters the month and rainfall data (findings), the *flowering intensity* and the *days to flowering* will be inferred. Rain IF: rainfall that induces a flowering, *Days to flowering*: days to flowering after Rain IF, Flowering: *Flowering intensity*.

2.4. Sensitivity analysis

A sensitivity analysis of a BN allows the user to quantify the changes in the values of a target variable due to findings (changes) in other model variables (Rositano et al., 2017). Depending on the variable type, there are specific metrics to use: variance reduction for continuous variables and mutual information (entropy reduction) for categorical variables. The higher the metric value, the higher the influence (Marcot, 2012; Norsys, 2023; Uusitalo, 2007). We ran a sensitivity analysis for our target variables, using mutual information for *flowering intensity* and variance reduction for *days to flowering* for each flowering event (F1-F4).

2.5. Validation procedure

Partitioned testing data was used to evaluate the model's performance in inferring *flowering intensity* and *days to flowering* over the four potential flowering events in the model (Marcot, 2017). The validation dataset included data for different years from four farms in the region (see Section 2.1); each flowering event was evaluated using the metric Spherical Payoff (SP); the SP scores range from 0 to 1, with 1 as the best performance (Marcot, 2012; Norsys, 2023).

3. Results and discussion

Here we introduce the first probabilistic flowering model based on long-term observed data capable of inferring multiple flowering events in a given crop year (Fig. 5). The model used few variables to infer flowering events and their intensities; the inference based on water availability (rainfall and months) and uses months as a proxy for the air temperature to estimate flowering events (Fig. 4). Previous studies on coffee flowering have mainly explored responses of flowering to changes in biophysical conditions within a single harvesting year. Alvim (1960) and Pagotto Ronchi and Rodrigues Miranda (2020) explored the effect of water availability on the anthesis; Masarirambi et al. (2009) the use of irrigation as a mechanism to concentrate the harvest; Drinnan and Menzel (1995) investigated the effect of temperature on flowering; and Lin (2008) examined the influence of microclimate changes on the flowering due to shading under coffee agroforestry systems.

3.1. Model sensitivity

Flowering intensity: Overall, the sensitivity analysis indicates that the other flowering events, the month, and the rainfall are the most influential – in this order (Fig. 6). Therefore, there is a strong influence among flowering intensities across the flowering events. Interestingly, for the *flowering intensity* for the first two flowering events (F1 and F2), the most influential variables are the FIs from the next flowering events (FI 2 and FI 3, correspondingly), and for the FIs for the last two flowering events are the FIs from the previous flowering events (FI 2 and FI 3, correspondingly). See Supplementary Material for the occurrence of flowering events reported in the region.

Days to flowering: Compared to *flowering intensity*, 1) the results

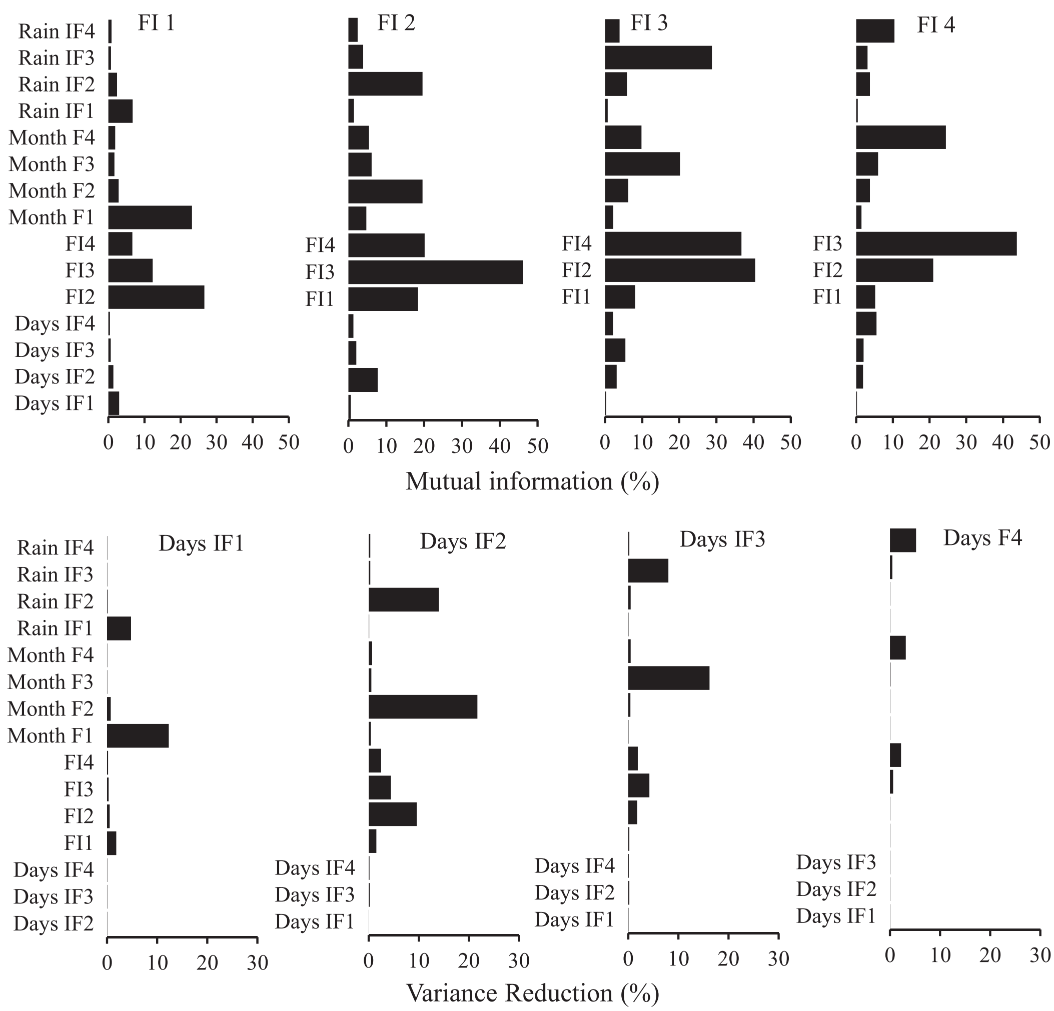

Fig. 6. Sensitivity analyses results for the flowering intensity (top) and *days to flowering* (bottom) for the four flowering events. Rain IF: rainfall that induces a flowering, Days IF: days to flowering after Rain IF, FI: *Flowering intensity*.

suggest that *days to flowering* are less sensitive to changes in the state of other variables, and 2) it is more a "local" variable for each flowering event because it does not have direct links to other flowering events. The month is most influential for the first three flowering events, followed by Rain IF, except for the fourth flowering event, where it is the opposite. The flower intensities have a lower influence on days to flowering.

The analysis screened the influence of all the node variables within the BN on a given target variable. Therefore, for a given flowering event, the analysis will also consider the influence of other flowering event variables, even those in the future. This information is relevant for backward and forward inference, where different queries can be placed to understand the flowering dynamic over the crop season. One practical use for sensitivity analyses is helping identify which variables should be prioritized for data collection, giving more effort to the variable(s) that impact more the target variable; in our case, rainfall is the only variable that would need measurements.

3.2. Model performance

The validation results indicate that the model performed well in inferring *flowering intensity*. The mean SP value over all farms was 0.78. At the farm level, the best performance in inferring *flowering intensity* was observed for the farm *San Francisco*, followed by *Jardin Botanico* and *San Jose*. Looking at the single flowering events (F), we can observe that the best performance of the model was for F2 and F3 at *San Francisco* and *San Jose*, and F1 and F4 at *El Rosal* and *Jardin Botanico* (Table 1). Unsurprisingly, the model performed best for *San Francisco*, as data on flowering and rainfall from this farm were used for the model training. In the case of the other farms, the possible variations in the actual rainfall (mm) they experienced versus the data registered in *Campos Azules* may partly explain the lower SP scores. Rainfall can have high spatial variability in the region, especially in the case of light rains, which means that the actual day and intensity of the rainfall on those three farms could be different from the one recorded in *Campos Azules*. In 2006, for example, from the 11 flowering events that occurred in total on the three farms, only two farms recorded a flowering event on the same day. Other factors may also play a role. On plantations with more than one canopy strata, rainfall interception can be so high that light rainfall may not reach the coffee plants at all or not in the same magnitude as in low-shaded or unshaded plantations (Lin, 2008; Siles et al., 2010).

The model did not infer the *days to flowering* following a rainfall event as well as the *flowering intensity*. The mean SP value over all farms was 0.45; even in San Francisco, the model only scored an SP value of 0.54. This may be linked to the low sensitivity the variable had toward other variables (see sensitivity analysis). These results suggest that the model may be further developed by adding variables or links to improve the prediction of *days to flowering*. We believe that a variable for temperature, such as the number of degree-days, would improve the performance (Montoya-Restrepo et al., 2009). However, during a meeting with a group of experienced coffee agronomists who work in the study region, we tested the model with them, and they agreed that the predicted trends aligned with their expectations. The improvements on *days to flowering* would thus be focused more on enhancing the precision than the accuracy.

3.3. Model testing

We ran the model with some basic scenarios to test if it would produce i) the expected logic outcomes or ii) inconsistencies with the existing knowledge or data (Marcot, 2017). First, we tested if the model would correctly infer that if there was no second flowering, the first and sole would be large, and there would not be a third and fourth flowering. Fig. 7 shows that if the state of the variable *Flowering 2* is set to "No" (by setting the finding of a probability of 100% that there is no second flowering), there is a 100% probability that the first flowering is "Large", and likewise a 100% probability that there will be no third or fourth flowering. The first flowering is most likely to occur in March (probability = 36%), or May (28%), and the rainfall preceding this flowering event would most likely be abundant and higher than 10 mm (44%). The most probable day for anthesis would be 10 ± 1.9 days after the rainfall that broke the water stress (Alvim, 1960).

Second, we tested if the model would correctly infer that having four flowering events in a single month is not possible. The model was able to match this expectation, but the certainty of the results was not the same for all months. If we selected January as the month the four potential flowerings would occur, the model inferred a 100% probability of a large first flowering and a 100% probability of no further flowerings occurring (Fig. 8a). If we chose March, there was a 63% probability of a large first flowering and a 54% probability of a second large flowering (Fig. 8b). And even though the model gave the highest probabilities of no further flowerings occurring (49% for F3 and 76% for F4), the chances of a third and even fourth flowering occurring were not zero.

Third, coffee flowerings early in the year (January–March) tend to be small, whereas later flowerings increase in intensity. We tested this by setting the dates for the first two flowerings to January and February and the dates for the third and fourth flowerings to May and June (Fig. 9). The BN inferred that there is a 94% probability that the first flowering will be small and a 79% probability that the second will be as well. The third flowering, set to occur in May, has an 81% probability of being large. Under this query, there is a high probability (68%) that no fourth flowering will occur; however, if the date of the fourth flowering is set to May instead of June, a large flowering will occur with a probability of 69% (not shown). The test shows that the inferred flowerings match our expectations about the date and corresponding size; this was also confirmed by the coffee agronomist during the model testing session mentioned above. See supplementary material for the occurrence of flowering events per month.

3.4. Learning from the model

In addition to providing the new insights described in the sensitivity analysis and model testing section, the model can also be used to learn the dynamics of coffee flowering and provide new insights into the functional relationship among variables (Hui et al., 2022). For example, exploring the dependencies between the rainfall inducing flowering,

Table 1
Spherical payoff values for *flowering intensity* and *days to flowering*.

Farms	Years (No.)	Flowering events (total)	Spherical payoff (SP)									
			Flowering intensity					Days to flowering				
			F1*	F2	F3	F4	Mean	F1	F2	F3	F4	Mean
San Francisco	5	14	0.82	1.00	0.90	0.78	0.88	0.63	0.57	0.77	0.20	0.54
San Jose	7	22	0.70	0.71	0.79	0.60	0.70	0.41	0.32	0.42	0.38	0.39
El Rosal	6	12	0.90	0.59	0.64	0.96	0.77	0.33	0.69	0.58	0.58	0.54
Jardin Botanico	6	12	0.99	0.51	0.61	0.97	0.77	0.31	0.23	0.00	0.71	0.31
Mean			0.85	0.70	0.74	0.83	0.78	0.42	0.45	0.44	0.47	0.45

* F1, F2: Flowering event 1, Flowering event 2, etc.

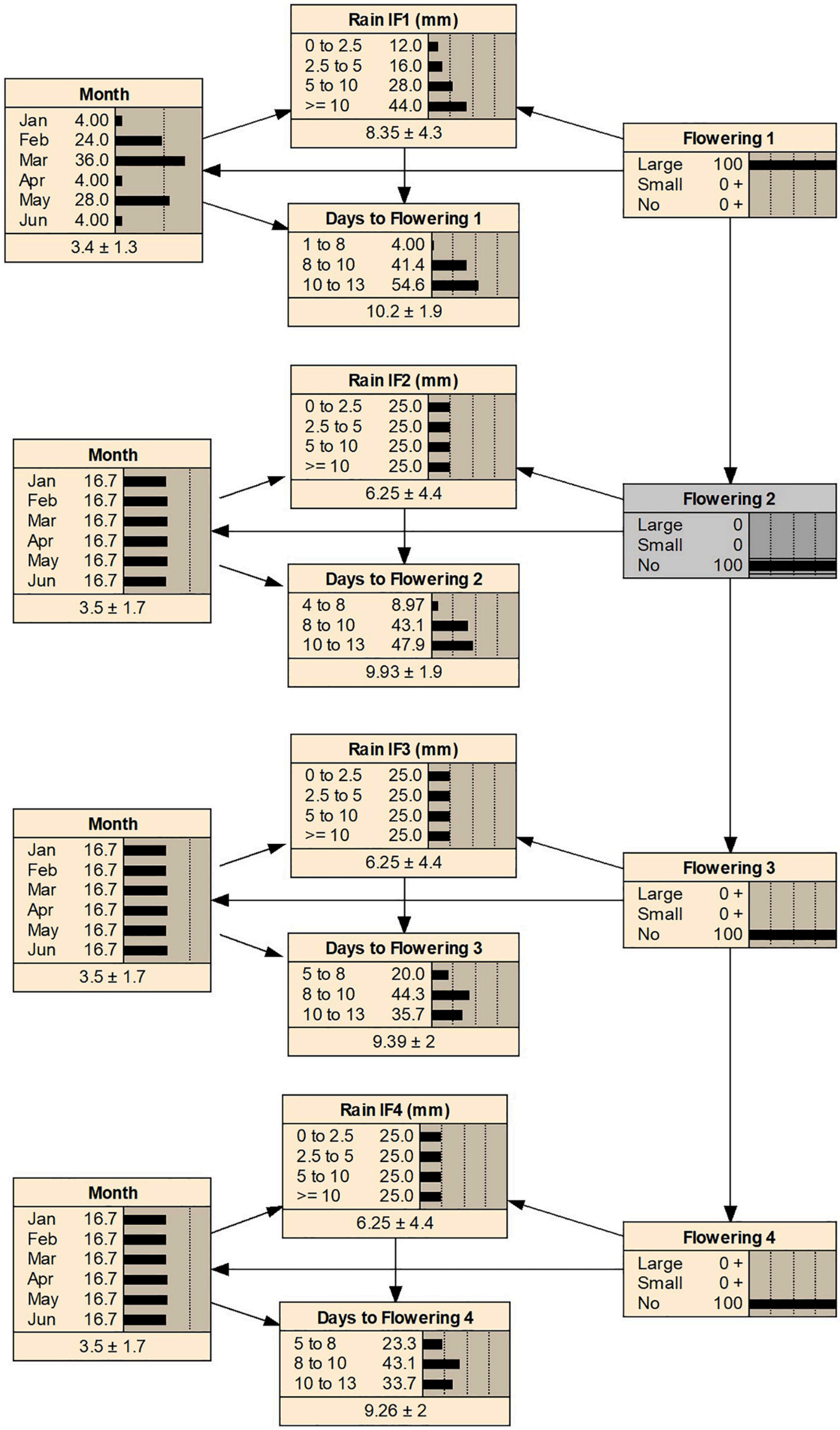

Fig. 7. Model testing case one: The model correctly infers that if there is no second flowering, there will also not be a third and fourth, and the only flowering occurring in the year is a large one. The grey box marks the node where the evidence was entered (finding).

month, and the flowering intensity for multiple flowering events, if we set the state of flowering intensity to large for any flowering event, the model will depict the most probable month(s) when this is possible and rainfall (mm) required for it (backward inference); which changes according to the flowering event (F1-F4). Because the model was trained with decades of data (prior information) of flowering events, the number of possible combinations of the variables' states is considerable, and Bayesian inference allows us to explore in detail the most probable outcomes in flowering if different conditions of month and rainfall occur for any flowering event (Table 2). Also, we can learn about combinations

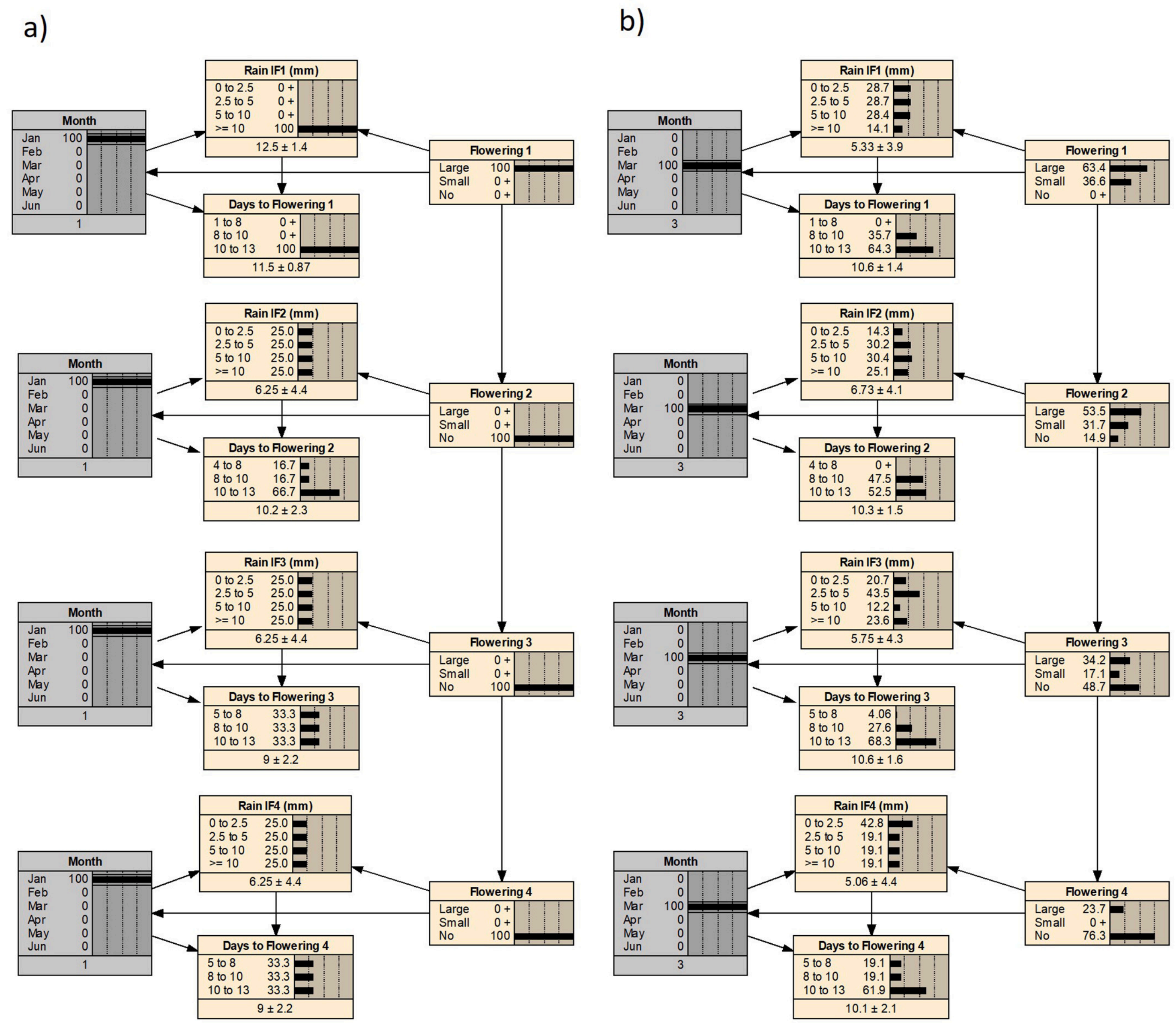

Fig. 8. Model testing case two: The model correctly infers that the probability of four flowerings occurring in a month is very low (a: all in January, b: all in March). The grey box marks the node where the evidence was entered (finding).

of variables states not observed; for example, light rains in May and April that provoke flowerings (Table 2).

Another example is the number of *days to flowering*. In January, February, or March, more days are required (10 to 13 days) between a rainfall and the resulting flowering; in April, May, and June, it takes fewer days (8 to 10 days). Furthermore, a large flowering event tends to occur more quickly (8 to 10 days) after rain than a small flowering (10 to 13 days).

In addition to learning from the model, the model can support farm planning or risk management processes. Knowing the details of an expected flowering event can help foresee potential yield losses due to heavy rainfalls during anthesis (Lara-Estrada et al., 2012; Murugan et al., 2022). If a farmer, for example, observes a rainfall event of over 10 mm in January, there is a 67% probability that the corresponding flowering event will be small. If, however, the farmer observes the same amount of rainfall in March, there is a 100% probability of a large flowering event. Assuming that heavy rains happen during anthesis for both cases, smaller yield losses are expected from the January flowering than from the one in March.

Finally, we did not present all the possible queries or new information on the multiple flowerings that can be learned from the model but expect that potential users will have the opportunity to. See the Supplementary Material for more examples of queries (e.g., model usage).

4. Future developments and applications

Even though the model can support decision-making and learning about flowering in its current form, further developments are planned. First, we plan to improve the inference for the number of *days to flowering* after a rainfall event so that the expected day of anthesis can be estimated as precisely as possible. This may be addressed by adding a temperature-based variable (e.g., mean temperature or degree-days) or additional links between variables; however, we still have not found temperature data matching the study region's flowering dataset (time and spatial resolution).

Second, we want to develop this coffee flowering model into a case for an index-based insurance against rainfall during coffee flowering (anthesis); for this, a yield loss component would be created and coupled

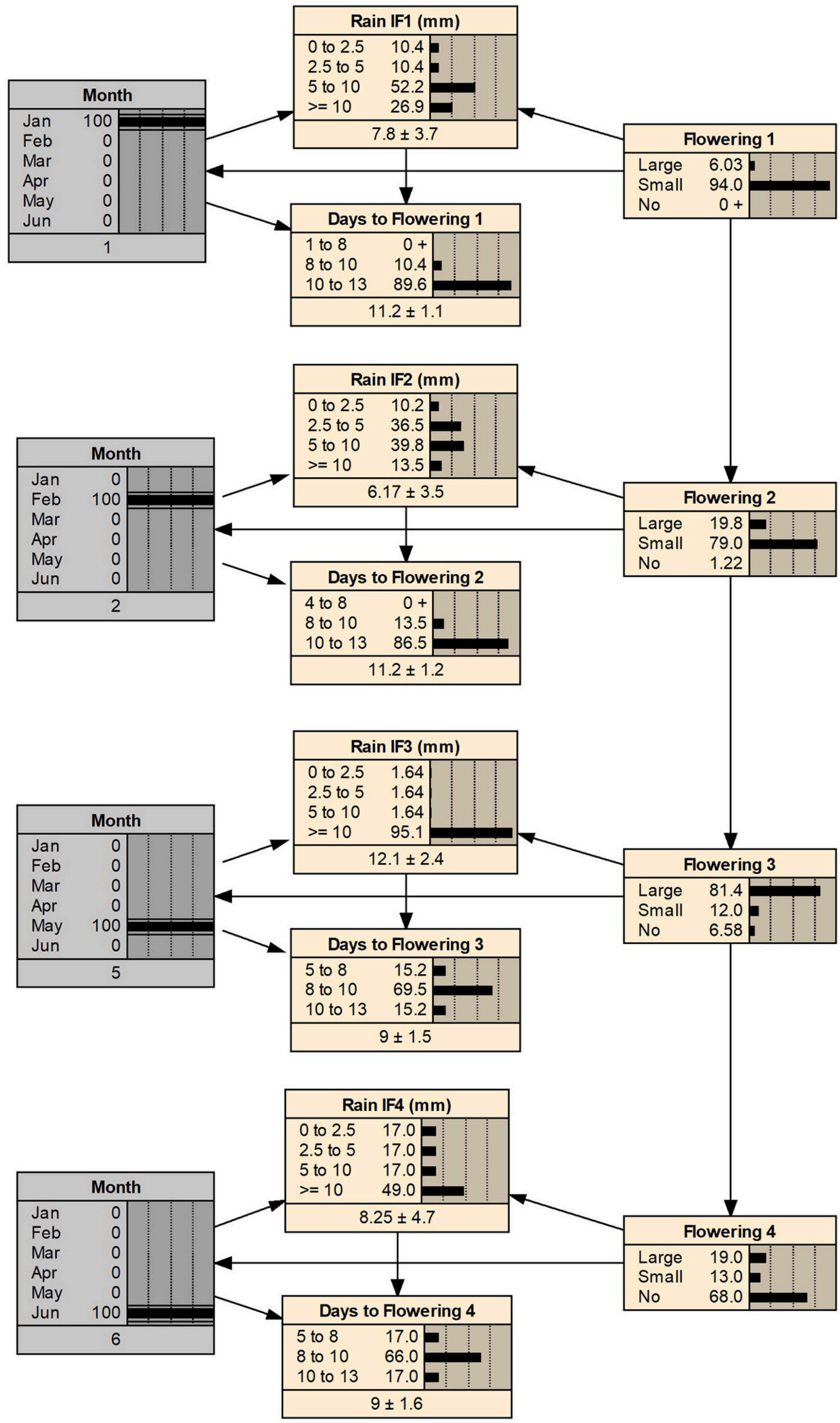

Fig. 9. Model testing case three: The BN correctly infers that flowerings occurring early in the year, January or February, are most likely small, whereas later flowerings (here May) have a higher probability of being large. The grey box marks the node where the evidence was entered (finding).

to the flowering model so that the resulting integrated model could infer the yield and economic losses due to rainfall during flowering and possible compensations. One study found that agricultural index-based insurance reduced the exposure to risk by an average of only 31% (Jensen et al., 2016). The risk reduction for insurance holders is higher in cases where the index used in the insurance is highly correlated with the insured risk, which requires the collection of loss data (Jensen et al., 2016). We have yearly yield data from the same farms reported in the study (longer than flowering data), from which we could estimate the yield losses; however, we will need to convert the flowering intensity from categorical (small, large, no) to continuous values (% of flower from the total possible flower per year: 100%), which is a challenge that would need to be solved. Once the model is developed, a validated Bayesian network model like this would offer a reliable, transparent, and easy-to-use option to reduce the basis risk for insurance holders by building finance resilience, and farmers would be able to take more

Table 2
Most probable flowering intensity results given rainfall in a given month for the first flowering (F1 in the model). Only the highest most probable outcome is displayed. Values in parentheses indicate probabilities.

Month	Rain IF1 (mm)			
	0 to 2.5	2.5 to 5	5 to 10	≥ 10
January	Small (100)	Small (100)	Small (100)	Small (67)
February	Small (100)	Small (80)	Small (50)	Small (60)
March	Small (50)	Small (50)	Large (75)	Large (100)
April	Small (50)	*	*	*
May	*	Large (100)	Small (50)	Large (100)
June	*	*	*	Large (100)

[*] Indicates that the corresponding level of Rain IF was not reported for the month, and a uniform distribution was produced for flowering intensity.

production risks and improve their livelihoods (Brouwer et al., 2014; Norton et al., 2014). It is likely that climate change and socio-economic development will further increase the demand for insurance coverage of the associated risks as increments in the mean daily precipitations are expected for March, April, and May in the study region (Clement et al., 2018; Imbach et al., 2018), which shows that further research and development in this field is of high priority.

5. Conclusions

In this paper, we introduce a flowering dynamic Bayesian network model that estimates the *flowering intensity* and *days to flowering* across multiple flowering events for *Coffee arabica* L. For this, the model uses only a small number of easy-to-measure predictor variables from coffee areas in the Pacific Region of Central America. The model was parameterized using five decades of observed data and can comprehensively depict the profile of flowering events in the Pacific coffee region of Nicaragua. It can also be used in other coastal Pacific coffee areas in Central America. We could show that the model has a good performance and can thus help farmers and other stakeholders to better understand the coffee flowering dynamics over a cropping season. Even though further work is required to improve the prediction of the number of *days to flowering*, we believe that the model in its current form represents a useful decision-support tool for coffee practitioners and scientists to forecast and monitor flowering events and help them plan the implementation of farming practices in coffee plantations. Furthermore, due to the simplicity and graphical structure of Bayesian networks, even individuals without a deep understanding of agricultural models can run scenario analyses and extract the results. Future potential use of this model in agricultural index-based insurance products for coffee farms will require further development and the inclusion of yield losses due to heavy rain during flowering.

CRediT authorship contribution statement

Leonel Lara-Estrada: Funding acquisition, Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing. **Luis Enrique Sucar:** Conceptualization, Methodology, Writing – review & editing. **Livia Rasche:** Visualization, Writing – original draft, Writing – review & editing.

Declaration of Competing Interest

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

Data availability

Data will be made available on request.

Acknowledgments

This work was supported by the Food and Nutrition Security Initiative (FaNSI) at the Natural Resources Institute, University of Greenwich, UK. Thanks to Dr Andrew Armitage for commenting on a previous version of the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ecoinf.2023.102434.

References

Aguilera, P.A., Fernández, A., Fernández, R., Rumí, R., Salmerón, A., 2011. Bayesian networks in environmental modelling. Environ. Model Softw. 26, 1376–1388. https://doi.org/10.1016/j.envsoft.2011.06.004.

Alvim, P.D.T., 1960. Moisture stress as a requirement for flowering of coffee. Science 132, 354. https://doi.org/10.1126/science.132.3423.354.

Anthony, F., Combes, M.C., Astorga, C., Bertrand, B., Graziosi, G., Lashermes, P., 2002. The origin of cultivated Coffea arabica L. varieties revealed by AFLP and SSR markers. Theor. Appl. Genet. 104, 894–900.

Bacon, C., 2005. Confronting the coffee crisis: can fair trade, organic, and specialty coffees reduce small-scale farmer vulnerability in Northern Nicaragua? World Dev. 33, 497–511. https://doi.org/10.1016/j.worlddev.2004.10.002.

Beall, K., Loisel, J., Medina-Cetina, Z., 2022. PermaBN: a Bayesian network framework to help predict permafrost thaw in the Arctic. Eco. Inform. 69, 101601 https://doi.org/10.1016/j.ecoinf.2022.101601.

Bertrand, B., Boulanger, R., Dussert, S., Ribeyre, F., Berthiot, L., Descroix, F., Joët, T., 2012. Climatic factors directly impact the volatile organic compound fingerprint in green Arabica coffee bean as well as coffee beverage quality. Food Chem. 135, 2575–2583. https://doi.org/10.1016/j.foodchem.2012.06.060.

Bornemisza, E., Collinet, J., Segura, A., 1999. Los suelos cafetaleros de América Central y su fertilización. In: Desafíos de la caficultura en Centroamérica. IICA, PROMECAFE, San José, C. R, pp. 97–137.

Bozzola, M., Charles, S., Ferretti, T., Gerakari, E., Manson, H., Rosser, N., von der Goltz, P., 2021. The Coffee Guide. International Trade Centre. https://doi.org/10.21256/zhaw-23941.

Brouwer, R., Tinh, B.D., Tuan, T.H., Magnussen, K., Navrud, S., 2014. Modeling demand for catastrophic flood risk insurance in coastal zones in Vietnam using choice experiments. Environ. Dev. Econ. 19, 228–249. https://doi.org/10.1017/S1355770X13000405.

Clement, K.Y., Wouter Botzen, W.J., Brouwer, R., Aerts, J.C.J.H., 2018. A global review of the impact of basis risk on the functioning of and demand for index insurance. Int. J. Disaster Risk Reduct. 28, 845–853. https://doi.org/10.1016/j.ijdrr.2018.01.001.

Descroix, F., Snoeck, J., 2004. Enviromental factors suitable for coffee cultivation. In: Wintgens, J.N. (Ed.), Coffee: Growing, Processing, Sustainable Production: A Guidebook for Growers, Processors, Traders, and Researchers. Wiley-VCH Verlag GmbH, Weinheim, Germany, pp. 164–177.

Drinnan, J.E., Menzel, C.M., 1995. Temperature affects vegetative growth and flowering of coffee (Coffea arabica L.). J. Horticult. Sci. 70, 25–34.

Eze, E., Girma, A., Zenebe, A.A., Zenebe, G., 2020. Feasible crop insurance indexes for drought risk management in northern Ethiopia. Int. J. Disaster Risk Reduct. 47, 101544 https://doi.org/10.1016/j.ijdrr.2020.101544.

Gay, C., Estrada, F., Conde, C., Eakin, H., Villers, L., 2006. Potential impacts of climate change on agriculture: a case of study of coffee production in Veracruz, Mexico. Clim. Chang. 79, 259–288. https://doi.org/10.1007/s10584-006-9066-x.

Goodyear, R., 2004. Irrigation. In: Coffee: Growing, Processing, Sustainable Production. John Wiley & Sons, Ltd, pp. 308–323. https://doi.org/10.1002/9783527619627.ch8d.

Hidalgo, H.G., Alfaro, E.J., Quesada-Montano, B., 2017. Observed (1970–1999) climate variability in Central America using a high-resolution meteorological dataset with implication to climate change studies. Clim. Chang. 141, 13–28. https://doi.org/10.1007/s10584-016-1786-y.

Hui, E., Stafford, R., Matthews, I.M., Smith, V.A., 2022. Bayesian networks as a novel tool to enhance interpretability and predictive power of ecological models. Eco. Inform. 68, 101539 https://doi.org/10.1016/j.ecoinf.2021.101539.
Imbach, P., Fung, E., Hannah, L., Navarro-Racines, C.E., Roubik, D.W., Ricketts, T.H., Harvey, C.A., Donatti, C.I., Läderach, P., Locatelli, B., Roehrdanz, P.R., 2017. Coupling of pollination services and coffee suitability under climate change. PNAS 201617940. https://doi.org/10.1073/pnas.1617940114.
Imbach, P., Chou, S.C., Lyra, A., Rodrigues, D., Rodriguez, D., Latinovic, D., Siqueira, G., Silva, A., Garofolo, L., Georgiou, S., 2018. Future climate change scenarios in Central America at high spatial resolution. PLoS One 13, e0193570. https://doi.org/10.1371/journal.pone.0193570.
Jensen, N.D., Barrett, C.B., Mude, A.G., 2016. Index insurance quality and basis risk: evidence from Northern Kenya. Am. J. Agric. Econ. 98, 1450–1469. https://doi.org/10.1093/ajae/aaw046.
Kath, J., Mittahalli Byrareddy, V., Mushtaq, S., Craparo, A., Porcel, M., 2021. Temperature and rainfall impacts on robusta coffee bean characteristics. Clim. Risk Manag. 32, 100281 https://doi.org/10.1016/j.crm.2021.100281.
Kögler, F., Söffker, D., 2017. Water (stress) models and deficit irrigation: system-theoretical description and causality mapping. Ecol. Model. 361, 135–156. https://doi.org/10.1016/j.ecolmodel.2017.07.031.
Lara-Estrada, L., Haggar, J.P., Ditmar, S., Bruno, R., 2012. Coffee yield variations and their relations to rainfall events in Nicaragua. In: Presented at the the 24th International Conference on Coffee Science, Association Scientifique Internationale du Café (ASIC), Costa Rica.
Lara-Estrada, L., Rasche, L., Schneider, U.A., 2017. Modeling land suitability for Coffea arabica L. in Central America. Environ. Model Softw. 95, 196–209. https://doi.org/10.1016/j.envsoft.2017.06.028.
Lara-Estrada, L., Rasche, L., Schneider, U.A., 2021. Land in Central America will become less suitable for coffee cultivation under climate change. Reg. Environ. Chang. 21, 88. https://doi.org/10.1007/s10113-021-01803-0.
Lara-Estrada, L., Rasche, L., Schneider, U.A., 2023. Exploring the cooling effect of shading for climate change adaptation in coffee areas. Clim. Risk Manag. 42, 100562 https://doi.org/10.1016/j.crm.2023.100562.
Lin, B.B., 2007. Agroforestry management as an adaptive strategy against potential microclimate extremes in coffee agriculture. Agric. For. Meteorol. 144, 85–94. https://doi.org/10.1016/j.agrformet.2006.12.009.
Lin, B., 2008. Microclimate effects on flowering success in coffee agroforestry systems. Am. Eurasian J. Agricult. Envirom. 3, 148–152.
Liu, K., Harrison, M.T., Hunt, J., Angessa, T.T., Meinke, H., Li, C., Tian, X., Zhou, M., 2020. Identifying optimal sowing and flowering periods for barley in Australia: a modelling approach. Agric. For. Meteorol. 282–283, 107871 https://doi.org/10.1016/j.agrformet.2019.107871.
Liu, K., Harrison, M.T., Yan, H., Liu, D.L., Meinke, H., Hoogenboom, G., Wang, B., Peng, B., Guan, K., Jaegermeyr, J., Wang, E., Zhang, F., Yin, X., Archontoulis, S., Nie, L., Badea, A., Man, J., Wallach, D., Zhao, J., Benjumea, A.B., Fahad, S., Tian, X., Wang, W., Tao, F., Zhang, Z., Rötter, R., Yuan, Y., Zhu, M., Dai, P., Nie, J., Yang, Y., Zhang, Y., Zhou, M., 2023. Silver lining to a climate crisis in multiple prospects for alleviating crop waterlogging under future climates. Nat. Commun. 14, 765. https://doi.org/10.1038/s41467-023-36129-4.
Marcot, B.G., 2012. Metrics for evaluating performance and uncertainty of Bayesian network models. Ecol. Model. 230, 50–62. https://doi.org/10.1016/j.ecolmodel.2012.01.013.
Marcot, B.G., 2017. Common quandaries and their practical solutions in Bayesian network modeling. Ecol. Model. 358, 1–9. https://doi.org/10.1016/j.ecolmodel.2017.05.011.
Masarirambi, M.T., Chingwara, V., Shongwe, V.D., 2009. The effect of irrigation on synchronization of coffee (Coffea arabica L.) flowering and berry ripening at Chipinge, Zimbabwe. In: Physics and Chemistry of the Earth, Parts A/B/C, 9th WaterNet/WARFSA/GWP-SA Symposium: Water and Sustainable Development for Improved Livelihoods, 34, pp. 786–789. https://doi.org/10.1016/j.pce.2009.06.013.
Montagnon, C., Marraccini, P., Bertrand, B., 2012. Breeding for coffee quality. In: Oberthür, T., Läderach, P., Cock, J.H. (Eds.), Specialty Coffee: Managing Quality. IPNI, Penang, Malaysia, pp. 89–117.
Montoya-Restrepo, E.C., Arcila-Pulgarín, J., Jaramillo-Robledo, A., Riaño-Herrera, N., Quiriga-Zea, F., 2009. Un modelo para simular la produccón potencial del cultivo del café en Colombia. Boletín Técnico - CENICAFE 33, 52.
Murugan, M., Alagupalamuthirsolai, M., Ashokkumar, K., Anandhi, A., Ravi, R., Rajangam, J., Dhanya, M.K., Krishnamurthy, K.S., 2022. Climate change scenarios, their impacts and implications on Indian cardamom-coffee hot spots; one of the two in the world. Front. Sustain. Food Syst. 6 https://doi.org/10.3389/fsufs.2022.1057617.
Muschler, R.G., 2001. Shade improves coffee quality in a sub-optimal coffee-zone of Costa Rica. Agrofor. Syst. 131–139 https://doi.org/10.1023/A:1010603320653.
Norsys, 2023. Netica Help [WWW Document]. Netica Help. URL https://www.norsys.com/WebHelp/NETICA.htm (accessed 7.1.22).
Norton, M., Osgood, D., Madajewicz, M., Holthaus, E., Peterson, N., Diro, R., Mullally, C., Teh, T.-L., Gebremichael, M., 2014. Evidence of demand for index insurance: experimental games and commercial transactions in Ethiopia. J. Dev. Stud. 50, 630–648. https://doi.org/10.1080/00220388.2014.887685.
Osorio, N., 2002. The Global Coffee Crisis: A Threat to Sustainable Development. International Coffee Organization, London, UK.
Pagotto Ronchi, C., Rodrigues Miranda, F., 2020. Flowering percentage in Arabica coffee crops depends on the water deficit level applied during the pre-flowering stage. Rev. Caatinga 33, 195–204. https://doi.org/10.1590/1983-21252020v33n121rc.
Rositano, F., Piñeiro, G., Bert, F.E., Ferraro, D.O., 2017. A comparison of two sensitivity analysis techniques based on four bayesian models representing ecosystem services provision in the argentine pampas. Eco. Inform. 41, 33–39. https://doi.org/10.1016/j.ecoinf.2017.07.005.
Schuch, U.K., Fuchigami, L.H., Nagao, M.A., 1992. Flowering, ethylene production, and ion leakage of coffee in response to water stress and gibberellic acid. J. Am. Soc. Hortic. Sci. 117, 158–163. https://doi.org/10.21273/JASHS.117.1.158.
Siles, P., Vaast, P., Dreyer, E., Harmand, J.-M., 2010. Rainfall partitioning into throughfall, stemflow and interception loss in a coffee (Coffea arabica L.) monoculture compared to an agroforestry system with Inga densiflora. J. Hydrol. 395, 39–48. https://doi.org/10.1016/j.jhydrol.2010.10.005.
Sucar, L.E., 2015a. Bayesian networks: Representation and inference. In: Probabilistic Graphical Models. Springer, London, London, pp. 101–136. https://doi.org/10.1007/978-1-4471-6699-3_7.
Sucar, L.E., 2015b. Dynamic and temporal Bayesian networks. In: Probabilistic Graphical Models. Springer, London, London, pp. 161–177. https://doi.org/10.1007/978-1-4471-6699-3_9.
Taylor, M.A., Alfaro, E.J., 2005. Central America and the Caribbean, climate of. In: Oliver, J. (Ed.), Encyclopedia of World Climatology. Encyclopedia of Earth Sciences Series. Springer, Dordrecht, Netherlands, pp. 183–189. https://doi.org/10.1007/1–4020-3266–8_37.
Uusitalo, L., 2007. Advantages and challenges of Bayesian networks in environmental modelling. Ecol. Model. 203, 312–318. https://doi.org/10.1016/j.ecolmodel.2006.11.033.
Uusitalo, L., Tomczak, M.T., Müller-Karulis, B., Putnis, I., Trifonova, N., Tucker, A., 2018. Hidden variables in a dynamic Bayesian network identify ecosystem level change. Eco. Inform. 45, 9–15. https://doi.org/10.1016/j.ecoinf.2018.03.003.
van Oijen, M., Dauzat, J., Harmand, J.-M., Lawson, G., Vaast, P., 2010. Coffee agroforestry systems in Central America: II. Development of a simple process-based model and preliminary results. Agrofor. Syst. 80, 361–378. https://doi.org/10.1007/s10457-010-9291-1.