Using Bayesian Networks to Model Flowering in Coffee Plantations in Central America

Leonel Lara-Estrada[1] **L. Enrique Sucar**[2]

[1]Natural Resources Institute, University of Greenwich, Medway Campus, Kent, UK
[2]Computer Science Department, Instituto Nacional de Astrofísica, Óptica y Electrónica, Puebla, Mexico

Abstract

The coffee plant is climate-sensitive; extreme rainfall during a flowering day causes reductions in coffee yields. The coffee flowering's intensity and date of occurrence are influenced by the water stress (dry season) and rainfalls during the dry and beginning of the wet season. Multiple flowering events could occur in a year, and there is a finite number of possible flowers per plant. This contribution introduces a Bayesian network model to infer multiple coffee flowerings for the Pacific Region in Nicaragua. The model structure and parameters were built based on previous related studies and data from coffee farms in the region, which included 55 years of flowerings (intensity and date of occurrence) and daily rainfall data from a coffee farm. Flowering data from 4 farms and rainfall data from two locations were used to validate the model using the metric spherical payoff: results were above satisfactory to infer flower intensity. Also, the model was able to depict expected phenological behaviors for single or multiple flowerings. We believe this model has the potential to evolve and support the development of an agricultural insurance to deal with yield losses because of extreme rainfall during flowering

1 INTRODUCTION

Coffee phenology is sensitive to extreme weather events. Heavy rains during a flowering day could produce losses of flowers and, therefore, reductions in the coffee yields, see Figure 1. Climate change is expected to change the climatic pattern and increase the occurrence of these extreme events in regions like Central America. Agricultural climate Index–Based Insurance (IBI) is an option to deal with extreme events by compensating farmers when certain extreme weather conditions occurred during a phenological crop phase. This type of insurance required modeling the affected phenological phase and the expected losses due to the extreme event. Therefore, to create a model for yield losses due to extreme rainfall during coffee flowering, we introduce a Bayesian network model to predict multiple coffee flowering for coffee areas in the Pacific Region of Nicaragua. The model was developed based on expert knowledge, and its parameters estimated from 55 years data from a farm in Nicaragua. An initial evaluation of the model on other nearby farms shows promising results for predicting flowering intensity.

2 METHODOLOGY

Multiple coffee flowering events of varied magnitude could occur during a crop year. The intensity and number of flowering events depend on the level of the accumulated water stress during the dry season and the amount of rainfall (mm) that temporarily relieves the water stress. So, the rainfall reactivates the coffee bud flowers in dormancy (because of the water stress), giving place to the flower anthesis some days later [Alvim, 1960].

2.1 MODEL BUILDING

The model was built and validated using related previous studies and data from coffee farms in the Pacific Region of Nicaragua [L. Lara-Estrada, 2012]. The data from coffee farms included flowering dates and flowering intensity (using a qualitative scale such as small and large). Registers of daily rainfall were obtained from some farms; however, most of them were incomplete, so, based on proximity and level of missing data, two series of rainfall were selected for training and validation. The selected variables were Flowering intensity (Small, Large, No), Month (From January to June), Rain that induces flowering ($0 - 2.5$ mm, $2.5 - 5$, $5 - 10$, ≥ 10), and Days to Flowering after Rain ($5 - 8$ days, $5 - 10$, $10 - 13$); see Figure 2.

Accepted for the 38th Conference on Uncertainty in Artificial Intelligence (UAI 2022).

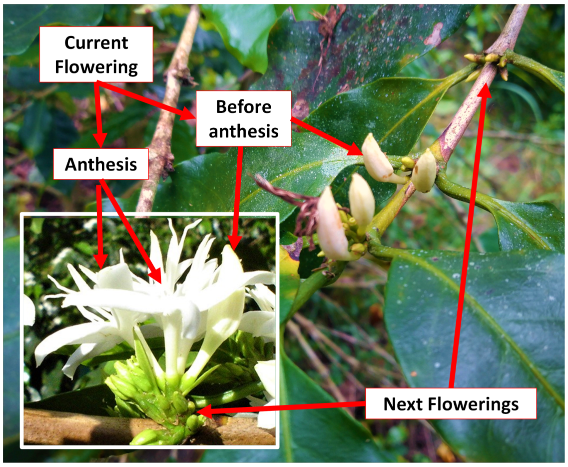

Figure 1: Coffee flowering buds. It can be seen flower buds that broke their dormancy on a day previous because of rainfall. Also, observe flower buds that continue in dormancy

The model estructure was built based on the flowering process described above (literature) and expert knowledge of the authors. Links between variables do not necessarily depict causal relationship; for example, instead of Month and Rain being the parents of Flowerings, they are children, so the only parent for any Flowering is the previous one; of course, except for Flowering 1. Overall, a given rainfall in a given month could trigger a flowering event of a given magnitude days after the rainfall. So, the model infers the flowering intensity and the date of occurrence using the month and rainfall (mm) as input information. According to the data, from one to four flowering events per year are the most common (90% of the cases), so only four flowerings are incorporated in the model.

The model, depicted in Figure 2, is composed of a base pattern representing the dependency relations between Month, Rain, Days to Flowering, and Flowering (intensity). This pattern is repeated for each flowering event four times to depict up to four flowering events. The links from one pattern or base structure to the next (similarly to dynamic Bayesian networks) are through the variable that represents the flowering intensity, as the number of flowering events is one of the factors that affect their intensity; for example, if there is only one flowering in the years it must be Large; for more examples see Figures 4 – 6.

2.2 MODEL TRAINING

The model was trained using 55 years of flowering and daily rainfall data (1943 – 1998) from the coffee farm San Francisco (San Marcos, Carazo); the flowering data include the day of flowering and intensity [L. Lara-Estrada, 2012]; see Figure 3. The variables days to flowering and rain that triggered the flowering were calculated based on the rains that occurred days after the registered flowering day. Due to the dataset did not include missing values, the Counting-Learning Algorithm [Norsys] was used to learn the model's conditional probability tables (CPTs).

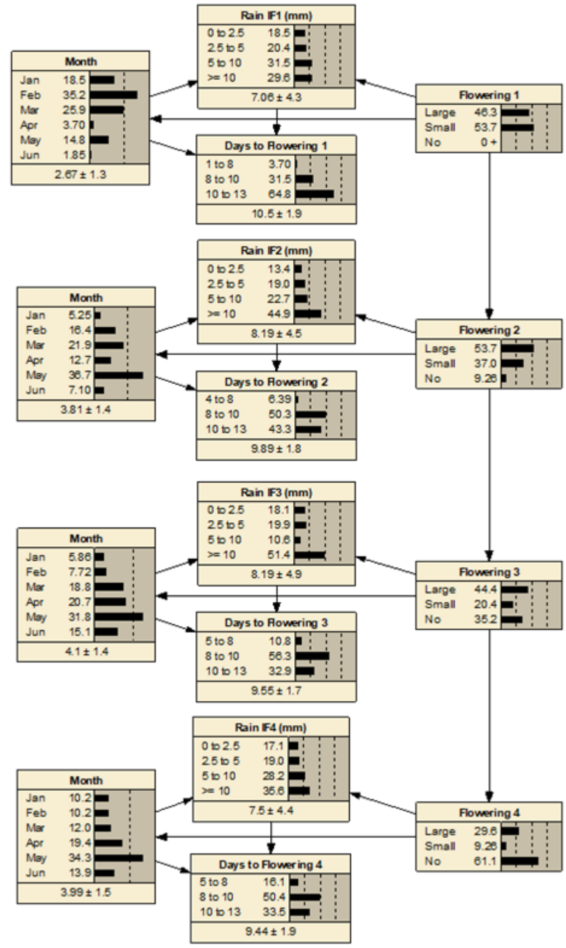

Figure 2: Coffee flowering model. A maximum of four flowering events are possible in this model. By entering the Month and Rain values, the intensity of Flowering and Days to Flowering will be inferred. Rain IF: Rain that induces flowering.

3 EXPERIMENTAL RESULTS

3.1 VALIDATION

It used the metric Spehrical Payoff (SP, score from 0 to 1, 1 is the best performance) to evaluate the model's performance for flowering intensity and days to flowering [Marcot, 2012], see Table 1. The flowering data used for validating the model were not included in the training and corresponded to three coffee farms: El Rosal (Jinotepe, Carazo), San Jose, and Jardin Botanico (Masatepe, Masaya), because of proximity ($5 - 7$ km). The rainfall data from the Research Station Campos Azules was used for the three farms. Therefore, it

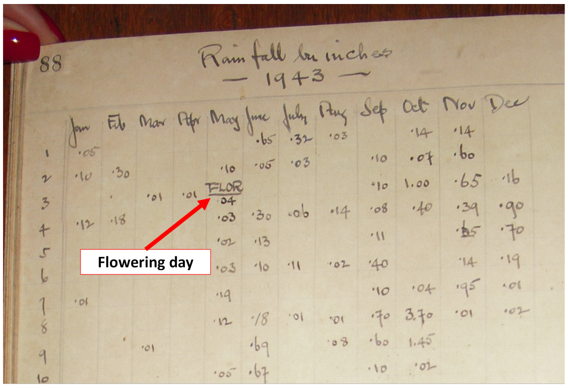

Figure 3: Example of Rainfall and coffee flowering registers obtained from the farm San Francisco (San Marcos, Nicaragua). “Flor” indicates a Large flowering and “Florcita” Small flowerings (not shown in the picture) in the registers. Data were registered in inches and converted to mm in this study.

used raining and flowering data from different farms than the ones used for training the model. It also included five years of flowerings and corresponding rainfall data (not used in the model training) from San Francisco [1].

For flowering intensity, the overall SP mean value for the farms evaluated was 0.78l; which is very good considering the distance between the El Rosal, San Jose and Jardin Botanico to Campos Azules, where the rainfall was registered. Rainfall has spatial variability, especially for light rains (observation from rainfall data in the region). So, the actual day and intensity of the rainfall on those three farms could be different. For example, in 2006, from 11 flowering events among the three farms, only two pairs of farms had one flowering event on the same day. At the farm level, the best performance was in San Francisco (expected), then Jardin Botanico, and San Jose. Looking at each flowering (F), we observed 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); however, the trend observed for San Francisco should be more reliable because flowering and rainfall data were from the same location. In the case of other farms, the possible variations in the actual rainfall (mm) they experienced versus the data registered in Campos Azules might partly explain the lower SP scores for those farms.

For the variable days to flowering, the model's performance was lower; the overall SP mean value was 0.45, even for San Francisco was 0.54, and much lower on the other farms. Even with this low performance, the overall tendencies align with the literature and expert comments, so the states for

Table 1: Spherical payoff values for Flowering intensity. F1 – Flowering 1.

Farms	Years (No.)	Flowerings (total)	Spherical Payoff F1 . F2 . F3 . F4
S. Francisco	5	14	0.82 1.00 0.90 0.78
S. Jose	7	22	0.70 0.71 0.79 0.60
El Rosal	6	12	0.90 0.59 0.64 0.96
Jar. Botanico	6	12	0.99 0.51 0.61 0.97

this variable need to be revised and explored if links to other variables need to be added to improve precision.

We also explored some basic scenarios, see Figures 4 – 6, and the model was able to match the expectations, including:

1. If there is no Flowering 2 (F2), there are no F3 and F4 (Figure 4).
2. If there is only one flowering (F1) during the year, it must be Large (Figure 4A).
3. It is not possible to have the four flowerings in a single month; less is possible. See Figure 5 for January and June.
4. Early flowerings in the year (January – March) tend to be of Small intensity (Figure 6A).
5. The occurrence and intensity of a flowering influence further flowerings, particularly the dependency between flowering events. Because there is a finite number of possible flowers per year, each flowering consumes a share of flowers that the following flowering events cannot use.

Also, the authors observed some interesting tendencies:

- Overall, flowerings from January to March take more days to occur after the induction rain.
- Independent of the month, Large flowerings take fewer days after the floral induction (Rain) than Small ones.
- Independent of the number of flowering, rainfalls below 2.5 mm tend to induce flowerings with Small intensity.
- Most of the flowerings occur before June, so it is less likely to be observed flowering in June.

3.2 USAGE AND FURTHER STEPS

Currently, the model has a very good performance in estimating the flowering intensity. However, improvements in the predictions for days to flowering will be required; this would potentially include adjustments in the variable states, adding a link to other variables, or adding other variables such as air temperature [J.E. Drinnan, 1995]. We believe that the simplicity and graphical structure of BNs would

[1] Years per farm: San Francisco: 1999, 2000, 2001, 2002, 2003; 2) San Jose: 2001, 2002, 2004, 2006, 2008, 2009, 2010; 3) El Rosal: 2005, 2006, 2007, 2008, 2009, 2010; and 4) Jardin Botanico: 2003, 2005, 2006, 2007, 2009, 2010.

make the model useful as a decision support tool suitable for coffee practitioners to foresee and monitor flowering events and help them to better plan the implementation of farming practices in the coffee plantations (e.g. agrochemical applications).

The evolution of this flowering model in a loss model for rainfall during flowering would support the usage of BNs as a potential methodology to develop and implement IBI. This type of insurance could become a financial adaptation strategy against expected climate change impacts for the region.

4 CONCLUSIONS

Knowing the flowering intensity and the days to flower makes a step forward in developing a crop loss model to infer possible losses in coffee yield due to extreme rainfall during flowering for the study region. Farmers and other studies have reported yield losses of about 60% because of rains during the flowering day. We believe the use of Bayesian Networks is a handy solution to depict a physiological process that would demand the use of more complex tools. We have developed a model for the conditions of the Pacific of Nicaragua that presents promising results for predicting flowering intensity.

Author Contributions

L. Lara conceived the initial idea, implemented the model, made the experimental evaluation and wrote a first draft of the paper. L. E Sucar contributed in the design of the model and revised the paper.

Acknowledgements

This work was supported by the Food and Nutrition Security Initiative funded by Research England's 'Expanding Excellence in England'. We thank the authors of [L. Lara-Estrada, 2012] for sharing their data on flowering and precipitation.

References

P.D.T. Alvim. Moisture Stress as a Requirement for Flowering of Coffee. *Science*, 132(3423):354–354, 1960.

C.M. Menzel J.E. Drinnan. Temperature affects vegetative growth and flowering of coffee (Coffea arabica L.). *Journal of Horticultural Science*, 70(1):25–34, 1995.

L. Lara-Estrada, J.P. Haggar., D. Stoian, B. Rapidel. Coffee yield variations and their relations to rainfall events in Nicaragua. In *International Conference on Coffee Science*, volume 24, 2012.

B.G. Marcot. Metrics for evaluating performance and uncertainty of Bayesian network models. *Ecological Modelling*, 230:50–62, 2012.

Norsys. Netica help. Available at `http://www.norsys.com/WebHelp/NETICA.htm` (2015).

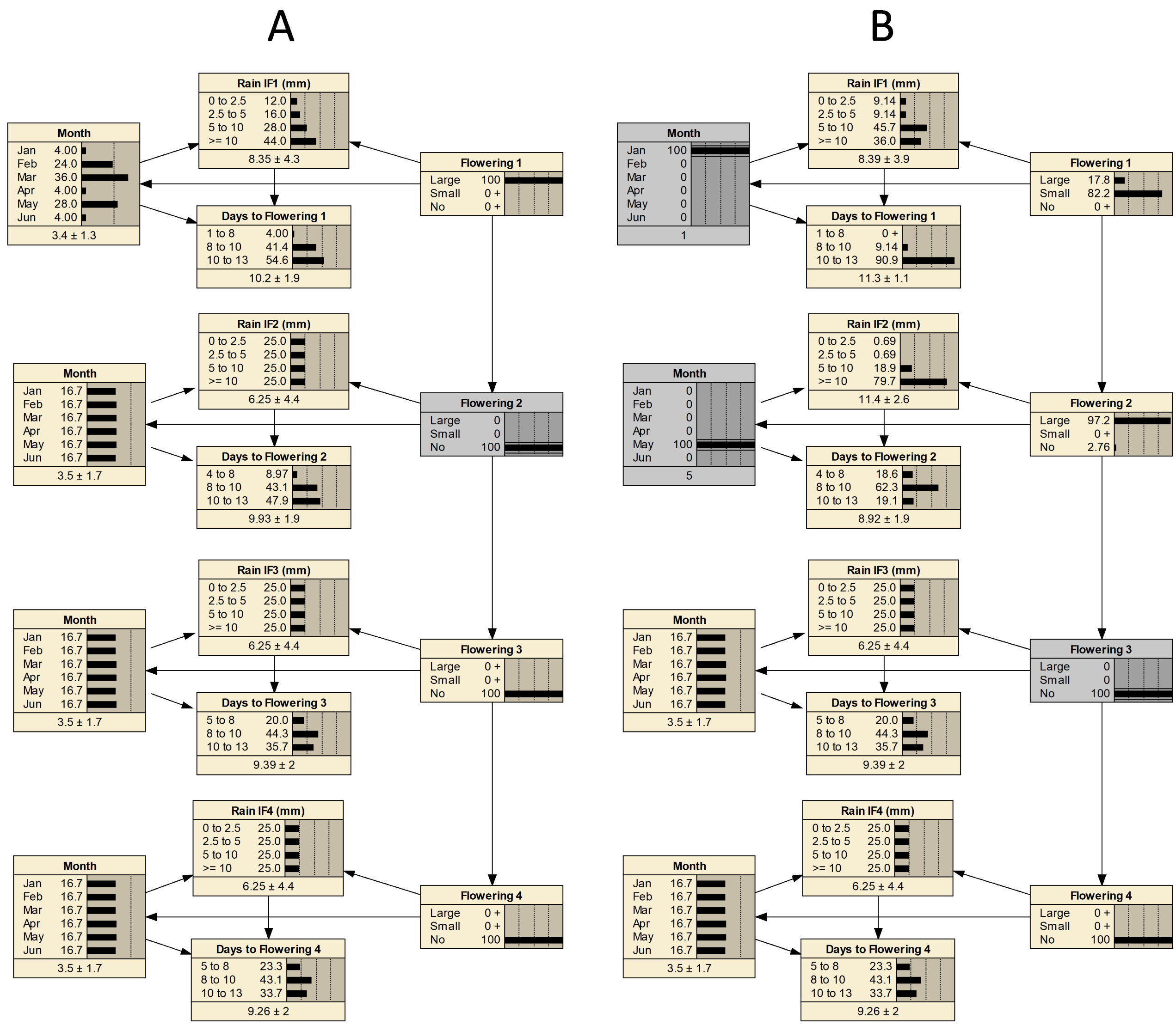

Figure 4: Testing the model with inference cases. A: What would be the most probable intensity and month of occurrence of one flowering in any year? B: What would be the most probable rainfalls and flowering intensities for two flowerings, one in January and one in May?.

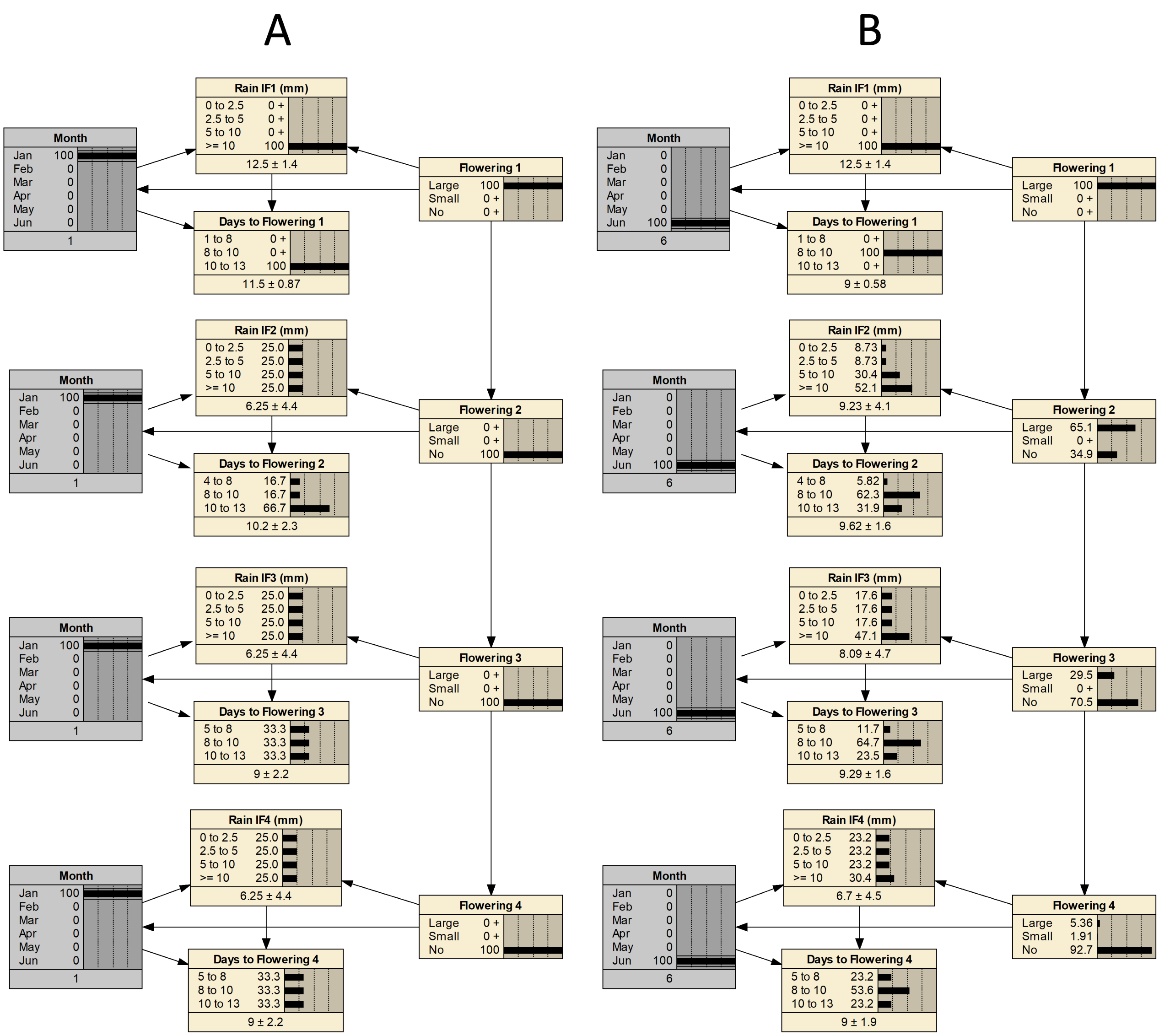

Figure 5: Testing the model with inference cases. Is it possible to have four flowering in January (A) or June (B)?

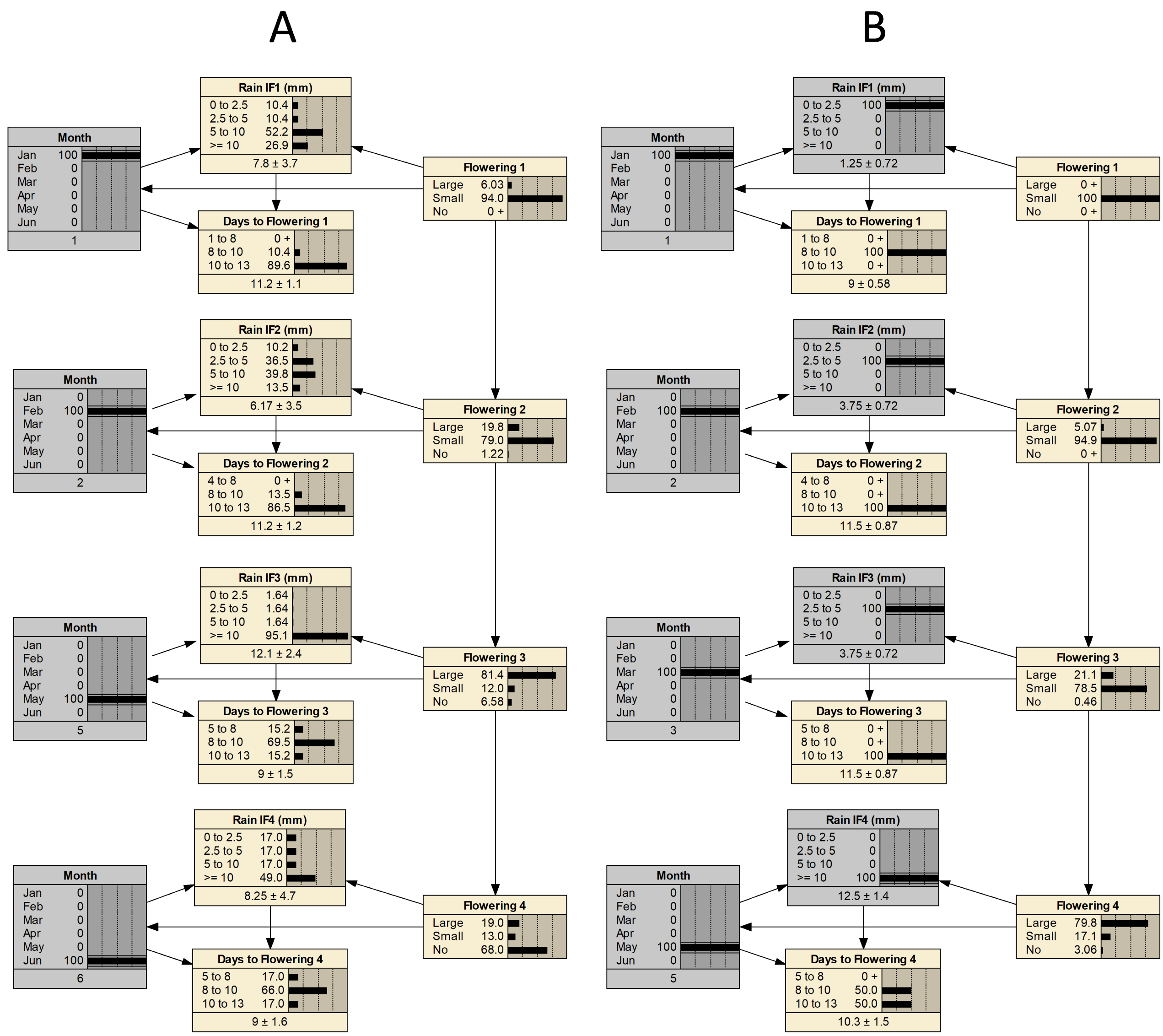

Figure 6: Testing the model with inference cases. A: What would be the flowering intensity if unknown rains occurred in January, February, May, and June? B: Inferring date and flowering intensity based on Month and Rain data.