The Design and Development of a Causal Bayesian Networks Model for the Explanation of Agricultural Supply Chains

MALLIKA KLIANGKHLO\textsuperscript{1}, SOMCHAI LIMSIROTRANA\textsuperscript{1}, AND BUKHOREE SAHOH\textsuperscript{2}

\textsuperscript{1}Department of Computer Engineering, Prince of Songkla University, Hat Yai, Songkla 90112, Thailand
\textsuperscript{2}Informatics Innovation Center of Excellence (IICE), School of Informatics, Walailak University, Tha Sala, Nakhon Si Thammarat 80160, Thailand

Corresponding author: Bukhoree Sahoh (bukhoree.sa@wu.ac.th)

This work was supported in part by the Prince of Songkla University, Thailand, through the Thailand Education Hub for ASEAN Scholarship in Doctoral Degree, under Grant TEH-233/2016.

ABSTRACT The balancing of demand and supply in the market is complex because of the dynamic supply chain and environment. It causes uncertain situations and is a limitation in decisions making systems that cannot produce reasonable descriptions to help decision makers eliminate uncertainties. This study proposes designing and developing a Causal Bayesian Networks (CBNs) model for market understanding, which encodes a human-like approach to explain demand and supply events for decision makers. A framework for generating reasonable descriptions in Agricultural Supply Chains (ASCs) management is proposed. The qualitative and quantitative design of the CBNs model is developed and proved that the CBNs model can reasonably explain events using predictive performance measurement and sensitivity analysis for producing reasonable descriptions. The results illustrate that the CBNs model is suitable for ASCs situation explanation involving uncertain situations and is ready to apply to real-world applications to support decision-making systems.

INDEX TERMS Explainable artificial intelligence, causal graph, machine learning, big data, demand and supply analysis, supply chain management.

I. INTRODUCTION

One of the problems in Agricultural Supply Chains (ASCs) is how to deal with the imbalance between demand and supply. For example, agricultural production and consumption may suffer from unexpected seasonal changes, such as an untimely harvest that causes shortages or even surplus market supply [1]. Decision making in this situation is a problem because there is a lack of comprehensive, real-time information.

The expansion of sensors-based-smart farming generates even more extensive data as a source of real-time information. It should allow decision makers to be more aware of demand and supply changes, and apply these variations to the benefit of supply chains. Unfortunately, big data suffer from the problems of enormous volume and complex dimensionality [2], [3]. Machine learning (ML) plays a vital role in a data-driven approach for supporting decision making [4]. It is widely used in agriculture decision-making systems because it can uncovered the information needed in ASCs [5]. For example, Punia et al. [6] has proposed a retail forecasting approach using extensive point-of-sales data, while Bu and Wang [7] utilized a water consumption approach for crop growth based on IoT sensors. They both employed deep learning to handle identification and classification, but it was less useful when decision makers wanted to ask how and why such outcomes were produced. Deep learning techniques produce black-box models which few people understand, and this lack of reasonable descriptions can cause decision makers to not fully understand the demand-supply situation, which may lead to poor decisions.

Reasonable descriptions utilize technologies that produce contextual information based on supply chain knowledge. The descriptions should be both testable and understandable by both human and agent-based systems by interpreting
supply chain knowledge using observational data. Fortunately, Bayesian Networks (BNs) for supply chain knowledge can produce reasonable descriptions since they determine transparent relationships using cause-and-effect as rational contextual information [8]. For example, Qazi et al. [1] and Ji et al. [9] employed BNs for managing a supply chain, by capturing relationships between supply chain factors from data based on correlation. However, they did not consider the causal assumptions based on rational human knowledge, which allows the model to detect irrational, unexpected ASCs events [10]. Our study addresses this drawback by proposing the use of Causal Bayesian Networks (CBNs) for knowledge in ASCs management.

CBNs determine the consequences and interdependencies among supply chain activities as a context synthesized from prior knowledge and big data. It models expert reasoning to explain demand and supply by producing reasonable descriptions.

The significant contributions of this study are:
- A new framework for knowledge description in ASCs management which addresses unexpected changes in supply chains.
- The development of a CBNs model for supplying descriptions in the natural rubber commodity market.
- Proof that the proposed model converges to expert reasoning by an analysis using predictive performance measurements and sensitivity analysis.

The rest of this paper is organized as follows: background knowledge and related work is presented in section II, and the descriptive supply chain management framework is introduced in section III. Section IV details the design and development of the CBNs model for supply chain management, and section V evaluates the CBN using quantitative experiments. Conclusions and future directions appear in section VI.

II. BACKGROUND KNOWLEDGE AND RELATED WORKS

ASCS management approach for supply chain in the presence of rapidly changing environmental conditions, and so is essential for efficient planning [11]. It utilizes real-time analysis and reaction, which depends upon contextual information in the supply chain [12]. In this section, we give some background on ASCs management, its common tasks, and various analysis approaches.

A. ASCs BACKGROUND

The futures market is an auction-based exchange where buyers and sellers trade contracts for deliveries set for a specified future date based on the quantity, quality, and price of commodities. The futures market helps ASCs protect their firm explains the market situation that can be in equilibrium and macro-level decisions that let stakeholders monitor these processes and determine the demand and supply at both levels [15]. Relationships between the supplier, customer, and firm explain the market situation that can be in equilibrium or exhibit abnormality as shortages and surplus. The relationship’s explanation helps decision makers to decide policies to deal with the supply chain abnormality, which shows how an ASC’s explanation so important for management.

B. ASCs MANAGEMENT APPROACH

Big data and ML have been employed for discovering demand and supply situations [16]. ASCs management uses the descriptive ability of ML to adjust operations and troubleshoot situations quickly and efficiently [17].

Kappelman and Sinha [18] proposed an approach based on stochastic optimization methods for dealing with uncertainty in supply chain systems. They claimed that their approach could effectively minimize uncertain problems and optimize time and complexity. Oh and Jeong [19] proposed a tactical supply planning model to overcome the short product life cycle and demand uncertainty. They concluded that their approach could provide the solution based on the optimal trade-off between profit and lead time. Gardas et al. [20] proposed systematic hierarchical structures using cause-and-effect-based relationships supporting decision-making. They discussed that their proposal could help decision-makers improve their understanding. The studies focused on maximizing profits in decision-making but did not consider explaining a situation of unbalancing between demand and supply.

BNs transparently model knowledge of supply chain relationships to produce such information [8], which decision-makers employ to create policies. BNs are probabilistic graphical models that can capture the uncertainty and relationships among relevant factors in the supply chain decision-making process. Random variables represent these factors, and their relationships are encoded by conditional probabilities using Bayes’ theorem.

Sharma [21], Chhimwal et al. [22], Lawrence et al. [23], and Ojha et al. [24] proposed for BNs-based risk assessment approach for supply chain management using historical data. They summarized that the approach could help the supply chain managers identify the risk factors early. El Amrani et al. [25] studied the sustainability of the supply chain network. These methods were successful because they focus on predicted outcomes and contextual explanations. However, they still did not consider explaining the context of demand and supply. This means that the model cannot answer ASCs management questions such as ‘What is the situation of demand and supply, and why were these outcomes
produced?’. The burden of causal interpretation and rational explanation is left to humans.

Causal Bayesian Networks (CBNs) have been proposed to address this problem [26], by modeling both emerging and rare events (e.g. climatic problems) that affect the management of ASCs. This means that CBNs will play an essential role for ASC explanations, even though the learning method for CBNs in ASCs is still far from decided.

C. CAUSAL BAYESIAN NETWORKS

CBNs are a human-like intelligent framework that encodes experience and knowledge based on cause-and-effect assumptions [27], [28]. In this way, CBNs extend traditional BNs by adding an interpretable ability in the manner of human-like understanding to produce explanations [26]. This lets CBNs explain demand and supply behavior to support ASC management.

A cause-and-effect assumption goes beyond correlation because it shows not only a statistical dependency between $X$ and $Y$, but encodes knowledge that $Y$ happens because of $X$. A CBNs assumption, $X \rightarrow Y$, is a cause-and-effect relationship that states that “Only $X$ can change $Y$”. For example, $\text{Weather} \rightarrow \text{Crop Yield}$ captures the idea that $\text{Weather}$ generally influences $\text{Crop Yield}$ which means $\text{Weather}$ is a cause of $\text{Crop Yield}$. A decision makers may ask “How will crop yield be undersupplied if prolonged rainfall is observed?”. This means that a $\text{Crop Yield} (\text{C})$ is denoted by undersupplied ($\text{us}$) given $\text{Weather} (\text{W})$ by prolonged rainfall ($\text{pr}$), then the query can be written using Bayes’ Theorem:

$$P(C = \text{us}|W = \text{pr}) = \frac{P(W = \text{pr}, C = \text{us}) \times P(C = \text{us})}{P(W = \text{pr})}$$

There are four probabilities in (1): the posterior $P(C = \text{us}|W = \text{pr})$; the likelihood $P(W = \text{pr}, C = \text{us})$, the prior $P(C = \text{us})$, and the observation $P(W = \text{pr})$. Their definitions are:

$P(C = \text{us}|W = \text{pr})$: the probability that an under supply is conditioned on prolonged rainfall;

$P(W = \text{pr}, C = \text{us})$: the likelihood that an under supply co-occur with prolonged rainfall;

$P(C = \text{us})$: the marginal likelihood of an under supply regardless of prolonged rainfall;

$P(W = \text{pr})$: the marginal likelihood of prolonged rainfall in the past.

This cause-and-effect assumption helps decision makers deal with market supply when influenced by the weather.

### TABLE 1. Causal structures.

| Causal Structure | Representation | Axiom |
|------------------|----------------|-------|
| Chain            | $X \rightarrow Z \rightarrow Y$ | $X$ indirectly causes $Y$ through $Z$ |
| Fork             | $X \leftarrow Z \rightarrow Y$ | $X$ and $Y$ are caused by $Z$ |
| Collider         | $X \rightarrow Z \leftarrow Y$ | $X$ and $Y$ are connected through $Z$ |

Although such observations are vital ingredients of CBNs for ACSs management, ACSs are a complex domain, which means that many causal assumptions cannot be expressed as direct $X \rightarrow Y$ relations, and may involve hidden factors between the $X$ and $Y$. Pearl et al. [29]’s causal model encodes such hidden factors—$Z$ based around three types of causal structures called chains, forks, and colliders. A chain encodes a cause-and-effect relationship in which the factor is involved sequentially. A fork encodes assumptions when a cause-and-effect relationship has a common cause. A collider encodes a cause-and-effect relationship which has a common effect. The structures are summarized in Table 1.

The causal structure uses conditional dependencies to connect nodes with causal relationships and block the paths between nodes with independencies; a process known as d-separation [30]. Causal discovery algorithms have been studies to structure a CBNs model from the observational data [31]. The algorithms are widely separated into two types: constraint-based and score-based. The constraint-based algorithms apply conditional independence constraints (e.g., Fast Causal Inference or FCI, and PC), while the score-based algorithms construct model using posterior probability of the candidate model (e.g. Greedy Equivalence Search or GES, and Greedy FCI). However, the resulted model’ performance is hard to be tested without a gold standard [32]. Then, expert-based modelling is the answer for discovering causal relationships in domain that lacks a baseline.

Causal relationships help ASCs model knowledge and help decision makers discover the reasons behind complex behaviors. The challenge is determining the semantics of the problem domains, verifying the dependencies among the random variables, and deciding whether they should connect or separate each other. This is the backbone of a descriptive supply chain management framework that can model micro-level and macro-level ASCs situations for decision-making.

### III. DESCRIPTIVE SUPPLY CHAIN MANAGEMENT FRAMEWORK

Useful supply chain management must produce proactive planning aligned with evidence, but this is not feasible with traditional technologies. This section proposes a new management framework that can generate explanations based on demand and supply evidence. The framework is summarized in Figure 1.

The framework consists of four components: data sensing, observation identification, situation explanation, and inference for making decisions; it follows the subdivisions employed by Belaud et al. [33].

Data sensing retrieves ASCs related data from sources such as global positioning systems (GPS), geographic information systems (GIS), remote sensing technologies, and web-based applications. The raw data is transformed into ASCs observations by the observation identification component. Although these observations detail ASCs information, they do not elaborate the relationships among the ASCs, which need deeper knowledge of the ASCs situation. The situation explanation produces rational explanations based on cause and effect in the manner of human-like reasoning. Lastly,
inference supports a decision-making component for proactive planning. It receives a hypothesis from a decision maker, and infers possible outcomes using the current ASCs situation. The resulting response and review help the decision maker to decide upon solutions and plan policy.

The intelligence of this framework depends upon the CBNs model developed as an initial requirement. This is the topic of the next section.

IV. DESIGN AND DEVELOPMENT OF CBNs MODEL FOR ASCs

CBNs are developed using causal discovery algorithms based on data dependencies that can structure the relations between random variables. Even the automatic algorithms, including Tree Augmented Naïve Bayes (TAN), Bayesian Network augmented Bayesian (BAN), and FCI, are widely studies [34], these algorithms generate statistical correlations among observations derived from well-structured and complete historical data that covers all possible events even it is the rarest. However, the unpredictability of the modern supply chain introduces uncertainty and change into the ASCs environment which generates rare events that do not exist in historical data. This means that a purely data-driven approach cannot produce accurate causal-and-effect explanations [35]. Moreover, the related data is still lacking in the context of the natural rubber ASCs. It lacks in both comprehensive and historical terms that is why the traditional ASCs runs with human.

To deal with that, using expert-based modeling as a gold standard. The prior knowledge depends upon the experts. It consisted of: (1) interviews with three experts, and two practitioners from the Central Rubber Market (CRM) in Hat Yai, Songkhla, Thailand; (2) reviews of a CRM database of 5 years provided by the Thai government. A prior-based process is required to integrate with data-driven process to produce a gold standard of CBNs model, as explained in the rest of this section.

A. RANDOM VARIABLES IN ASCs EXPLANATION

We employ random variables to model possible events, and quantify them based on observational data. Events are fixed as states but can occur randomly, according with natural change. We apply a traditional understanding of demand-supply price based on the structural representation employed in Pearl and Mackenzie [36]. Our case study models the futures market auction system as five ASCs explanation processes: source, supply, demand, market price, and futures market volatility. The random variables’ states are designed and built using prior knowledge from ASCs operations [5], and market price considers possible futures market conditions that contribute to our model.

We divide the random variables into three categories: observed, micro-level, and macro-level. Observed random variables model direct environmental observations, micro-level random variables represent supply chain activities, and macro-level random variables model market situations. The three categories of random variables and their states are detailed in Tables 2, 3, and 4.

1) OBSERVED RANDOM VARIABLES

The observed random variables are based on five ASCs explanations, which are summarized in Table 2.

**Climatic Problem** affects crop growth and harvesting, and can be obtained from weather station observations or open data services. **Plantation Area** estimates the crop yield quantity, which can be done manually or be automated using sensors. **Climatic Problem** and **Plantation Area** provide

| Context       | Random Variables         | States                      |
|---------------|--------------------------|-----------------------------|
| Source        | Climatic Problem Plantation Area | normal, drought, monsoon, flood downturn, sideways, uptrend, fluctuation |
| Supply        | Raw Material Cost Labor Resources | downturn, sideways, uptrend, fluctuation down, stable, up |
| Demand        | Exporting Costs Currency Exchanges | down, stable, up strengthening, stable, weakening |
| Future Market | Open Interest            | downturn, sideways, uptrend, fluctuation |
| Volatility    | Trading Volume            | downturn, sideways, uptrend, fluctuation |
| Market Price  | Future Market Prices Market Price | downturn, sideways, uptrend, fluctuation down, stable, up |

**TABLE 2.** Observed random variables.
information about the source and imply raw materials processing.

**Raw Material Cost**, such as crop price can be observed from open data services, and shows baseline information that harms secondary production. **Labor Resources** reflects production capacity, obtained through registered labor and official holiday figures. **Raw Material Cost** and **Labor Resources** are essential for estimating the supply context in the supply chain.

For demand, the required information relates to product consumption and logistics. **Exporting Costs** and **Currency Exchanges** movement are the critical factors. **Exporting Costs** information can be obtained from the petroleum prices index and **Currency Exchanges** from web services. In addition, production consumption is varied according to the agricultural product and the nature of the market. Some products may be traded through an agent, while many products are traded by auction, while the commodity product depends upon the futures market. This means that future market volatility is explained using **Open Interest**, **Trading Volume**, and **Futures Market Prices**, which can be observed from business data services. The **Market Price** is the index price for a commodity product reserved by a governmental office or agent, and is directly observable.

Although all the observed random variables in Table 2 are observable through open data, information systems, and services, we need to clarify the state of the variables for the specific market context. For example, in the case of the **Climatic Problem**, we focus on events such as **drought**, **monsoon**, and **flood** based on the vulnerability of the crop yield. The other variables are categorized based on movements (**down**, **stable**, **up**) and trends (**downtrend**, **sideways**, **uptrend**, **fluctuation**) derived from the ASCs non-stationary characteristics. The criteria for choosing a state is based on how its short term impact affects the trading process. For instance, **Raw Material Cost** shows the impact on manufacturing, while **Open Interest**, **Trading Volume**, and **Future Market Prices** reflect demand in the futures market. They are indirectly affected by trading processes in the long-term, and so their states are categorized based on trends.

### 2) MICRO-LEVEL RANDOM VARIABLES

The micro-level random variables are elaborated from prior knowledge of ASCs, which are summarized in Table 3.

**TABLE 3.** Micro-level random variables.

| Context          | Random Variables            | States         |
|------------------|-----------------------------|----------------|
| Source           | Crop Yield Producing        | low, normal, high |
| Supply           | Manufacturing Capacity       | low, normal, high |
| Demand           | Consumer Preference         | low, normal, high |
| Future Market    | Future Market               | down, stable, up |
| Volatility       | Movement                    |                |
| Source           | Crop Yield Producing        | low, normal, high |

**Crop Yield Producing** represents the level of market source. **Manufacturing Capacity** is the intermediate step of market supply production. **Consumer Preference** summarizes the product requirement, which reflects market demand. **Future Market Movement** is the external factor that influences market demand.

These micro-level random variables utilize **low**, **normal**, and **high** states which reflect their market context. However, **Future Market Movement** is defined using **down**, **stable**, and **up** values since it monitors the futures market situation.

### 3) MACRO-LEVEL RANDOM VARIABLES

Macro-level random variables summarize supply chain dynamics, which are represented using **ASCs Situation**. It consists of three possible states: **equilibrium** (the quantity demanded and supplied are the same), **shortage** (there is an excess of demand), and **surplus** (there is an excess of supply), but the relationships between demand, supply, and price are complex. For example, if demand is up and supply is down, then the **ASCs Situation** is a **shortage** that increases price according to market theory. In contrast, if the supply and demand relationship trigger a decreasing price, the **ASCs Situation** is still a **shortage** but with abnormal behavior. This latter scenario reflects a dysfunctional market policy, and market managers must implement corrections (i.e., by controlling the reference price or imposing a price ceiling). The states for the macro-level random variable are detailed in Table 4.

**TABLE 4.** The states of macro-level random variables.

| States          | Definition                                      |
|-----------------|-------------------------------------------------|
| equilibrium     | The market is equilibrium, and the price is stable. |
| abnormal-equilibrium | The market is equilibrium, but the price is rising or dropping. |
| shortage        | Excess demand and price is stable or rising.    |
| abnormal-shortage | Excess demand, but the price is dropping.  |
| surplus         | Excess supply and price is stable or dropping.  |
| abnormal-surplus | Excess supply, but the price is rising.      |

Table 4 lists the possible states for **ASCs Situation** in the context of demand, supply, and price. They are intended to help managers explain situations involving causal assumptions that interpret market behavior.

### B. ASSUMPTIONS IN CBNs MODELING

The futures market controls the demand of the natural rubber productions consumed by the automotive and tire industries [37], [38]. The products in that market are rubber sheets produced locally which depend on climatic conditions [39]. Indeed, climatic problems are the leading cause of decreased source production. We employ this information to model the causal assumptions between the random variables, and the resulting model is shown in Figure 2.

Figure 2 shows the graphical causal assumptions between the random variables, with a random variable for a cause pointing directly to effect random variable(s) (cause(s) →...
effect(s)). This graphical model can be interpreted into mathematical form using Structural Causal Model [29]. The assumptions are causally structured for explaining the ASCs Situation in terms of Manufacturing Capacity, Consumer Preference, and Market Price, and most of them are encoded as collides. For example, Trading Volume, Open Interest, and Future Market Price explain the liquidity and activity of Future Market Movement. Trading Volume reflects the short-term demanded quantity throughout the trading day, while Open Interest shows the number of futures contracts that are still open. Trading Volume and Open Interest are independent unless Future Market Movement is questioned, and then they become causally dependent. Crop Yield Production is also a collider, affected by Plantation Area and Climatic Problems. In other words, the causes are causally independent of each other, but conditioning on Crop Yield Production makes them dependent. Moreover, Crop Yield Production affects the behavior of Raw Material Cost, which passes its information to Manufacturing Capacity.

![FIGURE 2. Causal assumption of natural rubber SCs using CBNs.](image)

The CBNs model is initially constructed by casual assumptions as a rule-based prior knowledge. It is a qualitative knowledge that machine learning-based applications cannot reasonably interpret it. The data-driven approach is then employed to encode the causal assumptions to quantitative knowledge.

**C. DATA PREPROCESSING**

The character of ASCs-related data is multiple sources that become problematic because they are unstructured, redundant, and streaming. However, the real-world system must analyze and explain the event simultaneously and automatically for decision-making. For example, decision-makers may ask, “What does the trend of future market price look like given current evidence?”. The system must transform the input data from multiple sources into information represented with states of random variables to interpret and answer the question.

Section IV.A shows that states of random variables are discrete. For the observed random variable from the data source [40], Climatic Problem is categorized as normal, drought, monsoon, and flood. The rest are categorized by movement (down, stable, up) and trend (downtrend, sideways, uptrend, and fluctuation). Movement is the distance between points, and the trend is the semantic meaning of the direction of movements.

Time-series decomposition approach [41] is applied to produce the states that can extract information from multiple sources. This approach is dimensionality reduction that decomposes using a frequency-based interpolation function. Experts define the heuristics rules according to the short-long term impacts of each random variable on ASCs.

Natural rubber supply chains (SCs) data for tuning prior and likelihood functions were collected between 2015 and 2019 from the CRM in Hat Yai, Songkhla, Thailand. We randomly split the data collection into two subsets. The first subset is utilized for model training and validation, and the second is for model evaluation. The data-splitting method was performed using the scikit-learn Python library [43].

**D. MODELING CAUSAL ASSUMPTIONS**

CBNs are a qualitative model based on causal assumptions extracted from background knowledge, which are quantified using observational evidence by training their parameters.

The natural rubber SCs obtained from the training data are summarized in Table 5.

**TABLE 5. Summarization of natural rubber SCs.**

| Data Sources                  | Random Variables  | States               |
|-------------------------------|-------------------|----------------------|
| Climatological Center [42]   | Climatic Problem  | normal (48%), drought (8%), monsoon (13%), flood (31%) |
| Agricultural Production Data [43] | Labor Resources          | down (10%), stable (81%), up (9%) |
| Plantation Area             | Raw Material Cost   | downrend (36%), sideways (7%), uptrend (47%), fluctuation (9%) |
| Thailand Daily Rubber Price [44] | Market Price        | down (19%), stable (69%), up (19%) |
| Bank of Thailand [45]       | Currency Exchanges  | strengthening (53%), stable (4%), weakening (43%) |
| Markets Insider [46]        | Exporting Costs     | down (7%), stable (75%), up (19%) |
| Tokyo Commodity Exchange (TOCOM) [47] | Trading Volume | downtrend (47%), sideways (6%), uptrend (47%), fluctuation (0%) |
|                             | Future Market Price | downtrend (24%), sideways (11%), uptrend (34%), fluctuation (31%) |
|                             | Open Interest       | downtrend (47%), sideways (15%), uptrend (51%), fluctuation (12%) |

Table 5 summarizes random variables whose states are distributed and chose to show the movements that affect the market. However, we did not consider Plantation Area factors since rubber trees must grow for seven years before their first harvest and live for two decades. This means that
they remain stable and less significant during the 5-year data collection period used here.

While, micro and macro-level random variables are contextual variables, retrieved from CRM database. They are labelled using experts, shown in Table 6.

Table 6 shows micro-level and macro-level variables’ prior distribution. The major proportion of Crop Yield Producing is up (48%), which causes Manufacturing Capacity to be high (52%), which accounts for over half of the dataset. This suggests that the supply situation for this ASCs market has always been high. In contrast, the Future Market Movement value up (15%) is the lowest event occurrence, so cannot boost market demand, which results in Consumer Preference being normal (52%). This shows that the supply and demand situation is unbalanced, causing ASCs Situation to have an equilibrium value of 7%.

| Random Variables       | States                                |
|------------------------|---------------------------------------|
| Crop Yield Producing   | down (39%), stable (14%), up (48%)    |
| Manufacturing Capacity | low (17%), normal (31%), high (52%)   |
| Consumer Preference    | low (9%), normal (52%), high (39%)    |
| Future Market Movement | down (33%), stable (51%), up (15%)    |
| ASCs Situation         | equilibrium (7%), abnormal-equilibrium (29%), shortage (13%), abnormal-shortage (8%), surplus (24%), abnormal-surplus (20%) |

The states in Table 5 and Table 6 become the priors of the random variables. For example, let \( c_p \) be a set of \( m \)-possible outcomes of Climatic Problem(CP), and \( P(CP) \) be the prior for Climatic Problem, defined as: \( P(CP = normal) = 0.48 \), \( P(CP = drought) = 0.08 \), \( P(CP = monsoon) = 0.13 \), and \( P(CP = flood) = 0.31 \), according to the Climatic Problem entry in Table 5. This prior informs us that between 2015 and 2019, Thailand suffered from floods and monsoons almost half the time (which fits with typical tropical climate characteristics). Also, the priors of the states are not equally likely because of the nature of the Climatic Problem.

This data can be utilized to tune the likelihood parameters, according to Bayes’ Theorem. For example, the causal assumption shows that Crop Yield Producing (CYP) is affected by Climatic Problem (CP). The likelihood can be calculated using joint probability of this causal assumption, represented using a Conditional Probability Distribution (CPD). We use Maximum Likelihood Estimation [48] for tuning the likelihood parameters. For example, the CPD of Crop Yield Producing given Climatic Problem in the natural rubber supply chain is summarized in Table 7.

The CPD of Crop Yield Producing given Climatic Problem shows the low production is affected from strange weather (i.e., probabilities of low in Crop Yield Producing are 0.982 and 0.995 given drought and flood respectively). CPD show the likelihood between cause and effect random variables that is required for Bayes’ Theorem to explore posterior in CBNS model.

Although the CBNS model is encoded from expertise knowledge that makes human sense, it needs model validation to measure its performance for machine understanding.

### E. CBNS MODEL VALIDATION

The causal structure represents the scientific assumptions underpinning the ASCs data, and the CBNS-based model exhibits predictive ability with reasonable explanations. The purpose of CBNS model validation is to confirm that our proposed model can predict the ASCs situation.

According to the ASCs Situation’s states in Table 6, the target class is distributed over 6 possible outcomes and is imbalanced. Marcot and Hanea [49] proposed that 10-fold is the optimal value for k-fold cross-validation for a discrete Bayesian-based model. It resamples the data into ten subsets, using nine subsets in each iteration for training, and the rest for testing. Therefore, we have also employed 10-fold cross-validation to estimate model performance.

The metric for interpreting validation results is accuracy, selected by measuring the model’s predictive performance during the learning process; the results are shown in Table 8.

| ASCs situation 6-possible outcome | Accuracy (k = 10) |
|-----------------------------------|-------------------|
| equilibrium                       | 0.86              |
| abnormal-equilibrium              | 0.97              |
| shortage                          | 0.92              |
| abnormal-shortage                 | 0.88              |
| surplus                           | 0.96              |
| abnormal-surplus                  | 0.95              |
| Average                           | 0.94              |

Table 8 shows that the overall performance is high of 94%. The accuracies of equilibrium and abnormal-shortage are lower than the others because they are rare events, occurring at around 7% and 8% in the sample proportion. The equilibrium market is ideal and rarely occurs because the market context changes dynamically. Similarly, abnormal-shortage means a shortage of supply with decreasing price, which is an extraordinary situation that contradicts the laws of demand and supply. It is also a rare event with a small sample for training the model.

The validation shows that our proposed possesses good model performance and can be applied to this case study. Although k-fold cross-validation is fundamental for model
TABLE 9. Predictive performance comparison.

|                | Equilibrium | Abnormal-equilibrium | Shortage | Abnormal-shortage | Surplus | Abnormal-surplus | Avg. |
|----------------|-------------|----------------------|----------|------------------|--------|-----------------|------|
|                | PS RC FM PS RC FM PS RC FM PS RC FM PS RC FM PS RC FM |              |          |                  |        |                 |      |
| NN             | 0.87 0.83 0.85 0.96 0.95 0.96 0.93 0.92 0.93 | 0.91 0.88 0.89 0.93 0.95 0.94 0.93 0.95 0.94 | 0.93      |
| SVM            | 0.87 0.83 0.85 0.97 0.95 0.96 0.95 0.92 0.94 | 0.95 0.88 0.91 0.95 0.94 0.92 0.92 0.96 0.94 | 0.94      |
| DT             | 0.87 0.86 0.87 0.96 0.97 0.96 0.95 0.92 0.93 | 0.95 0.88 0.91 0.94 0.96 0.95 0.95 0.95 0.95 | 0.94      |
| NB             | 0.84 0.64 0.72 0.94 0.70 0.81 0.95 0.90 0.93 | 0.95 0.87 0.90 0.95 0.87 0.90 0.76 0.94 0.84 | 0.84      |
| BS             | 0.88 0.85 0.87 0.95 0.98 0.96 0.96 0.93 0.95 | 0.96 0.89 0.92 0.94 0.96 0.95 0.96 0.95 0.96 | 0.93      |
| CBNs           | 0.86 0.87 0.86 0.96 0.97 0.97 0.95 0.92 0.94 | 0.96 0.88 0.92 0.94 0.96 0.95 0.96 0.95 0.96 | 0.95      |

V. RESULTS

This section evaluates how well our CBNs model can perform the task correctly and rationally. Consequently, our experiments have two parts: 1) tests of the predictive performance for model correctness, and 2) sensitivity analysis for model reasonableness.

A. PREDICTIVE PERFORMANCE MEASUREMENT

This experiment measures the CBNs model’s predictive performance. The target class are the states of the ASCs Situation random variable since it helps to provide final decisions in the supply-chain system.

The states of ASCs Situation were measured based on Precision, Recall, and F-Measure. Precision (PS) is a proportion of the correction of the positive prediction, which is computed as $PS = \frac{TP}{TP+FP}$. TP is a true positive prediction, and FP a false positive prediction. Recall (RC) is a proportion of the correction of the prediction, which is computed as $RC = \frac{TP}{TP+FN}$. FN is a false negative prediction. F-Measure (FM) is a balance between Precision and Recall, which is computed as $FM = \frac{2 \times Precision \times Recall}{Precision + Recall}$.

The measurements employed the dataset described in Section IV.E. Baselines model were evaluated using testing dataset. The average scores of each model are shown in Table 9.

We used standard classification algorithms to compare the predictive performance of our proposed model, including geometric-based models (i.e., Neural Networks (NN) and Support Vector Machines (SVM)), logic-based models (i.e., Decision Trees (DT)), and probabilistic-based models (i.e., Naïve Bayes (NB), Bayesian Search (BS)). As we know that the performance of the classifiers depends upon algorithm’s parameters. Then, these models are implemented by scikit-learn [50], a Python library, with default parameter setting. For example, NN was set with 100 hidden layers, 0.001 learning rate, 200 epochs, ReLU as the activation function, and adam as the optimization. The comparative performance of the predicted results with our CBNs model highlights its predictive ability.

Table 9 shows the Precision, Recall, and F-measure for each model based on the states for ASCs Situation states. The average results were well over 80%, which is acceptable for prediction systems. The lowest was 84% from the NB model, since it employs a “naïve” assumption that its features are independent and only depend on the outcomes. This is not true for supply chains where features typically do rely on each other. The other results for NN, SVM, DT, BS, and CBN were 93%, 94%, 94%, 93%, and 95% respectively, which are high since all the models were trained and validated using well-prepared data. This suggests that these models are ready to apply to decision support systems to help understand the ASCs Situation.

The FM scores for the Equilibrium state are the lowest since it is an infrequent event that is sensitive to the balance of demand and supply, which is affected in various ways. It is also an ideal event, with little chance of occurring, but decision makers need to understand all the factors that support their decisions.

B. SENSITIVITY ANALYSIS FOR CAUSAL ASSUMPTION

Even though the CBN model’s results can be acceptably applied to prediction systems, it does not give an explanation for supporting decision making. We addressed this by conducting a sensitivity analysis to show the strength and sensible of connections between random variables. This show how well the CBNs provide guarantees on the query results with rationale explanations. Crucially, this aspect of the CBNs model is missing from the other models. The CBNs model provides contexts for supporting decision making in term of the state parameters of the random variables that impact ASCs Situation.

The BS-based model was compared with our model because of its use of conditional dependency of a Bayesian Network [51], which produces relationships based on a DAG of data dependency. We applied scenario-based sensitivity analysis to highlight the rational explanation of both models.

As a base case, we used the most sensitive scenario, “ASCs Situation is equilibrium”. That event has the lowest probability of occurring, but has the highest impact on decision making. According to our hypothesis, the posterior probabilities
of ASCs Situation may be affected by Manufacturing Capacity, Consumer Preference, and Market Price, and we assume that the base case is sensitive to variations of the states from its random variables.

Sensitivity analysis calculates posterior probability distributions based on the partial derivative over the unknown variables; for example, ASCs Situation is questioned given evidence (i.e., the evidence is a state of a cause random variable, e). It can be calculated as 
\[
\text{Sensitivity}(x) = \frac{\partial p(x|e)}{\partial e} 
\]

is a target variable, with interest in \(x = x_t\) as a base case, and \(p(x_t|e)\) posterior distribution of the base case given evidence. This computation is based on an algorithm contributed by Kjaerulff and van der Gaag [52].

The average sensitivity conditioned from all evidence is between zero and one. Zero means the changes of the ASCs Situation’s causes reduce the absolute change in the posterior probability of the base case that shows robustness in posterior distribution calculation, while one makes ASCs Situation more likely to occur. Sensitivity analysis can measure a minor change in the sensitivity of the ASCs Situation’s posteriors (i.e., the causes of non-equilibrium). In essence, this analysis computes the sensitivity between cause and effect in the manner of expert reasoning based on the uncertainty of the ASCs Situation.

The degree of sensitivity between causes and effects are represented using tornado diagrams since they are easy to read and interpret [53]. The x-axis in Figure 3 represents the sensitive of “ASCs Situation = equilibrium” between zero and one. The y-axis-bar lists the set of parameters as conditions that affect equilibrium. The random variables states have 27 possible values, but only the five of the most sensitive parameters appear in Figure 3.

Figure 3 shows the sensitivity levels for the base case from the CBNs and BS models. The sensitive degrees for CBNs and BS are 0.069 and 0.071 respectively.

One difference between CBNs and BS is the number of random variables affecting the sensitivity of the base case. CBNs is highly sensitive to Market Price, Manufacturing Capacity, and Consumer Preference, while BS is highly sensitive to Market Price, Manufacturing Capacity, Consumer Preference, Trading Volume, and ASCs Situation. The number of variables reflects upon resources and processing time.

The first three parameters from the models show that equilibrium has converged to zero. It means that changes to Manufacturing Capacity, Consumer Preference, and Market Price cause ASCs Situation to become unbalanced (=equilibrium, shortage, or surplus). The posterior distributions of ASCs Situation for both CBNs and BS are highly sensitive to Market Price. Experts understand that consumer and supplier behaviors are principal factors affecting ASCs Situation, and so BS and CBNs can help people interpret events using something close to expert reasoning.

The last two parameters in the CBNs and BS models are different for ASCs Situation. The CBNs is highly sensitive to demand (Consumer Preference), supply (Manufacturer Capacity), and price (Market Price), but the BS model is sensitive to Trading Volume because the training data may provide high correlations which lets the BS connect it, contradicting human understanding. Indeed, this relationship is considered an irrational explanation because Trading Volume is never used to explain ASCs Situation. Experts understand that Trading Volume is the root cause of ASCs Situation that transfers its effect through Future Market, Exporting Costs, and Consumer Preference. This is the situation for the CBNs which show that the equilibrium state of the ASCs Situation is sensitive to changes in Manufacturing Capacity, Consumer Preference, and Market Price. The sensitivity represents how domain experts view environment changes, and what they should consider adjusting first.

C. DISCUSSION

The experiments show that CBNs provide predicted outcomes and also relevant parameters to help decision makers understand the ASCs situation.

The first experiment confirms that the CBNs model has satisfactory performance in a market situation. CBNs can reach an accuracy of around 95%, which works well within traditional supply chain management, where many companies employ experts to examine the probabilities of shortage or surplus. However, small companies lack this expertise, which makes their analysis much more labor-intensive and time-consuming.
Recent experiment of model performance employs basic models and the FM of Equilibrium is 0.86, which is quite low. In the future, we plan to use a dynamic CBNs to improve our model performance and may compare with advanced model, such as random forest, gradient boosting, and deep learning.

The second experiment shows that CBNs offer a new dimension of decision support for supply chain management. It provides market interpretable explanations based on cause-and-effect, which is needed by companies.

VI. CONCLUSION

This study has proposed a Causal Bayesian Networks (CBNs) model for supporting market understanding. It produces reasonable explanations to aid decision makers dealing with ASC demand and supply uncertainty by interpreting contextual information based on big observational data.

We compared standard machine learning models (Naïve Bayes, Neural Networks, Support Vector Machines, Decision Trees, and Bayesian Search) to our CBNs model. Their performances for predicting unknown events were over 90%, but our model could reach around 95%. Sensitivity analysis confirmed that the CBNs model could produce reasonable descriptions of expert reasoning and that the model was sensitive to contexts utilized by experts. Our model can help decision makers better understand agricultural supply chain situations and successfully adjust supply chain mechanisms.

However, CBNs based on expert knowledge have a subjective quality, which means that markets with different supply chain characteristics will need to adjust the CBNs’ causal assumptions, and re-tune parameters with different historical data. In future work, we will examine other market elements, discover additional causal assumptions, and address the issue of exponential numbers of relevant random variables. The ongoing ASCs will grow continuously and modernly and the FM of Equilibrium is 0.86, which is quite low. Sensitivity analysis confirmed that the CBNs model could produce reasonable descriptions of expert reasoning and that the model was sensitive to contexts utilized by experts. Our model can help decision makers better understand agricultural supply chain situations and successfully adjust supply chain mechanisms.

In future work, we will examine other market elements, discover additional causal assumptions, and address the issue of exponential numbers of relevant random variables. The ongoing ASCs will grow continuously and modernly and generate many data covering rich and exciting events for further ASCs management. More data create more factors and opportunities to run ASCs management with better performance, which may be beyond the labor work in expert-based modeling. We then hope to perform studying on automatic algorithms to learn more inclusive knowledge from the modern ASCs and compare it with our proposed gold standard. We plan to employ causal discovery algorithms to determine causal relationships in the CBNs model to reduce labor-intensive and time-consuming tasks.

REFERENCES

[1] A. Qazi, A. Dickson, J. Quigley, and B. Gaudenzi, “Supply chain risk network management: A Bayesian belief network and expected utility based approach for managing supply chain risks,” Int. J. Prod. Econ., vol. 196, pp. 24–42, Feb. 2018, doi: 10.1016/j.ijpe.2017.11.008.

[2] Y. Huang, Z.-X. Chen, T. Yu, X.-Z. Huang, and X.-F. Gu, “Agricultural remote sensing big data: Management and applications,” J. Integrative Agric., vol. 17, no. 9, pp. 1915–1931, 2018, doi: 10.1007/s11393-017-6189-8.

[3] S. Wolpert, L. Ge, C. Verduou, and M.-J. Bogaardt, “Big data in smart farming—A review,” Agric. Syst., vol. 153, pp. 69–80, May 2017, doi: 10.1016/j.agsy.2017.01.023.

[4] R. Toorajipour, V. Sohrabpour, A. Nazarpour, P. Oghazi, and M. Fischl, “Artificial intelligence in supply chain management: A systematic literature review,” J. Bus. Res., vol. 122, pp. 502–517, Jan. 2021, doi: 10.1016/j. jbussres.2020.09.009.

[5] R. Sharma, S. S. Kamble, A. Gunasekaran, V. Kumar, and A. Kumar, “A systematically literature review on machine learning applications for sustainable agriculture supply chain performance,” Comput. Oper. Res., vol. 119, Jul. 2020, Art. no. 104926, doi: 10.1016/j.cor.2020.104926.

[6] S. Punia, S. P. Singh, and J. K. Madaan, “A cross-temporal hierarchical framework and deep learning for supply chain forecasting,” Comput. Ind. Eng., vol. 149, Nov. 2020, Art. no.106796, doi: 10.1016/j.cie.2020.106796.

[7] F. Bu and X. Wang, “A smart agriculture IoT system based on deep reinforcement learning,” Future Gener. Comput. Syst., vol. 99, pp. 500–507, Oct. 2019, doi: 10.1016/j.future.2019.04.041.

[8] S. Hosseini and D. Ivanov, “Bayesian networks for supply chain risk, resilience and ripple effect analysis: A literature review,” Exp. Syst. Appl., vol. 161, Dec. 2020, Art. no.113649, doi: 10.1016/j.eswa.2020.113649.

[9] G. Ji, L. Hu, and K. H. Tan, “A study on decision-making of food supply chain based on big data,” J. Syst. Sci. Syst. Eng., vol. 26, no. 2, pp. 183–198, Apr. 2017, doi: 10.1007/s11424-016-9303-3.

[10] E. Bareinboim and J. Pearl, “On pearl’s hierarchy and the foundations of causal inference,” Causal AI Lab, Columbia Univ., Tech. Rep. R-60, 2020.

[11] X. Zhao, R. Lovreglio, and D. Nilsso, “Modelling and interpreting pre-evacuation decision-making using machine learning,” Autom. Constr., vol. 113, May 2020, Art. no.103140, doi: 10.1016/j.autcon.2020.103140.

[12] P. Danese, M. Molinaro, and P. Romano, “Investigating fit in supply chain integration: A systematic literature review on context, practices, performance links,” J. Purchasing Supply Manag., vol. 26, no. 5, Dec. 2020, Art. no.100634, doi: 10.1016/j.pursup.2020.100634.

[13] L. Zhou and J. Huang, “Contagion of future-level sentiment in Chinese agricultural futures markets,” Pacific-Basin Finance J., vol. 61, Jun. 2020, Art. no. 101316, doi: 10.1016/j.pacificbaj.2020.101316.

[14] S. Chopra and P. Meindl, Supply Chain Management: Strategy, Planning, and Operation, 5th ed. London, U.K.: Pearson Education, 2013.

[15] M. J. Murray-Prior, M. F. Rois-Rubén, and P. O. Hartley, “Micro and macro-level approaches to modelling decision making,” Agricult. Syst., vol. 69, nos. 1–2, pp. 63–83, Jul. 2001, doi: 10.1016/S0308-521X(01)00018-X.

[16] Y. Zhan and K. H. Tan, “An analytic infrastructure for harvesting big data to enhance supply chain performance,” Eur. J. Oper. Res., vol. 281, no. 3, pp. 539–574, Mar. 2020, doi: 10.1016/j.ejor.2018.09.018.

[17] The Future Supply Chain for an Unpredictable World. IBM Inst. Bus. Value, Armonk, NY, USA, 2020.

[18] A. C. Kappelman and A. K. Sinha, “Optimal control in dynamic food supply chains using big data,” Comput. Oper. Res., vol. 126, Feb. 2021, Art. no. 105117, doi: 10.1016/j.cor.2020.105117.

[19] J. Oh and B. Jeong, “Tactical supply planning in smart manufacturing supply chain,” Robot. Comput.-Integr. Manuf., vol. 55, pp. 217–233, Feb. 2019, doi: 10.1016/j.cim.2018.04.003.

[20] B. B. Gardas, R. D. Raut, N. Cheikhrouhou, and B. E. Narkhede, “A hybrid decision support system for analyzing challeges of the agricultural supply chain,” Sustain. Prod. Consumption, vol. 18, pp. 19–32, Apr. 2019, doi: 10.1016/j.spc.2018.11.007.

[21] S. K. Sharma and S. Sharma, “Developing a Bayesian network model for supply chain risk assessment,” Supply Chain Forum, Int. J., vol. 16, no. 4, pp. 50–72, Jan. 2015, doi: 10.1080/16258312.2015.11728693.

[22] M. P. Shinwala, S. Agrawal, and G. Kumar, “Measuring circular supply chain risk: A Bayesian network methodology,” Sustainibility, vol. 13, no. 15, pp. 1–22, 2021, doi: 10.3390/su13158448.

[23] J.-M. Lawrence, N. U. I. Hossain, R. Jaradat, and M. Hamilton, “Leveraging a Bayesian network approach to model and analyze supplier vulnerability to severe weather risk: A case study of the U.S. pharmaceutical supply chain following hurricane Maria,” Int. J. Disaster Risk Reduction, vol. 49, Oct. 2020, Art. no.101607, doi: 10.1016/jijdrr.2020.101607.

[24] R. Olha, A. Ghadge, M. K. Tiwari, and U. S. Bititci, “Bayesian network modelling for supply chain risk propagation,” Int. J. Prod. Res., vol. 56, no. 17, pp. 5795–5819, Sep. 2018, doi: 10.1080/00207543.2018.14674059.

[25] S. El Amrani, N. U. I. Hossain, S. Karam, R. Jaradat, F. Nur, M. A. Hamilton, and J. Ma, “Modelling and assessing sustainability of a supply chain network leveraging multi echelon Bayesian network,” J. Cleaner Prod., vol. 302, Jun. 2021, Art. no.126855, doi: 10.1016/j.jclepro.2021.126855.
