CHAPTER 2

Literature Reviews

2.1. Longan in Season Production System

The majority of longan trees cultivated by Chiang Mai farmers are of Daw or E-Daw, (Figure 1) which is an early variety. The main advantage of this variety is that it blooms and sprouts fruit earlier than any other variety. These trees can be harvested before other variety in the area. Thus, it is highly praised and sought after by foreign markets. The shape of fruit is round and large with a big seed. The flesh is white and yellow, sweet and mild. It also has distinctive aroma. Its growth is quick and bountiful. It is sold as both fresh and preserved fruit such as dried or canned (Manochai et al., 2004).

Figure 1 Longan, Daw or E-Daw variety that is widely grown in Chiang Mai Province.

In general, the longan grower work in their orchards all year round, for example to fertilize, water, and control pests. They prop the branches up with bamboo supports to provide wind protection especially from December to February. The trees blossom in early January onto February. The harvesting time begins in July with the season coming to a close in August. The workflow of longan production during one season is shown in Figure 2.

Activities	Jan.	Feb.	Mar.	Apr.	May	Jun.	Jul.	Aug.	Sept.	Oct.	Nov.	Dec.
Watering, fertilization, pest control and weed control	←	—	—	—	—	—	—	—	—	—	—	→
Supporting the branches with bamboo	←	→										←→
Caring during flowering period	←→											
Harvesting and marketing fruits							←	→				
Pruning									←→			

Figure 2 Workflow of longan production within a 1-year period.

Source: interviews

2.2 Factors Affecting Longan Performance

According to the Department of Agriculture (2013) there are many factors related to longan blossom and fruit producing. Longan fruit production takes 6-8 months from the start of blossom to harvest time. The tree requires a lot of nutrients especially in the prolific year. Some nutrients in soil are taken away with fruits it produces. This later depletes the soil nutrients. Therefore, the soil requires fertilizers for nutrient replenishment in time. Moreover, only 4-5 months of fertilizer feeding time is not enough for longan. If it is not taken care well enough, the tree will not be sufficiently fertilized. If the environment is not suitable, the next-year yield could be less or it could not provide blossom. The most important factor is the apical growth that strongly relates to the environment. During November to December it is the time before longan blossom. The tree might stop producing new leaves and shoots until the winter comes, which stimulate a high value of blossoms. The environment factors affecting longan blossoming are:

1) Temperature: temperature is the important factor for longan regarding bud producing. The temperature suitable for longan to flower is 10-12 degrees celcius which normally appears in winter during November - December. This low temperature for a period of time can stimulate longan flowers to be blossomed.

2) Sunlight: longan usually produces flowers at the apical zone which is most exposed to sunlight. It would produce lesser amount of flowers at shaded position. Longan growth rate is influenced by duration of sunlight. It related to the amount of Carbohydrate inside the trunk. Longan flowers when total nonstructural carbohydrate (TNC) in its trunk is at a high level. TNC level will decrease when the longan blossoms and continue decreasing until harvesting time. According to the analytical results, TNC rates will be higher in the year before longan has extraordinary blossom than the year before it has less flower blossom.

3) Soil humidity: if soil humidity shows low level and the weather is cool before flowers blossom, longan tree will be stimulated to produce better flowers. The degree of soil humidity will affect the water absorption ability of a plant and will decrease the nitrogen amount that a plant can absorb. If the plant can absorb more nitrogen, the rate of producing leaves would be higher while the rate of producing flowers would be lower. Furthermore, low soil humidity could affect Abscisic acid hormones that subsequently reduce longan growth rate. With low growth rate, longan will have higher opportunity for producing flowers.

4) Fertilizer application: the amount of minerals from fertilizer that longan need should be added with appropriate quantity and at suitable time. The right amount of minerals from fertilizer could stimulate the production of flowers. Minerals can enhance plant's health. For trees with similar age, the more fertilized longan trees would earlier produce flowers. Normal watering and fertilization should be operated particularly at the time the longan tree is producing flowers and fruits (Department of Agriculture, 2002).

2.3 Tools for Crop Yield Estimation

It was found that there were several popular tools, as follows; crop modeling and mathematics modeling in the framework of crop yield estimation or prediction. The details are as follows;

2.3.1 Crop Simulation Model

The usage of statistical information, pertinent to cultivation such as meteorological information, agronomic principles, and knowledge base, is relevant for understand the farming and physiology of the crop. This pertinent information is gathered for relateing to all factors that affect the growth of the plants within that model. Each factor may also contain a knowledge or Mathematics Model within it. This models were applied to annual crops such as CERES-Rice model for rice (Timsina and Humphreys, 2006), CANEGRO model for sugarcane (Knox et al., 2010), CERES-Maize model for maize (Timsina and Humphreys, 2006), and the MunThai model for cassava. So far this kind of model has not yet been applied to perennial plants, which have a longer period of production.

2.3.2 Regression Modeling

This is a mathematical model, which shows the relationship between a set of yield data and other interesting variables that explain to and benefit the growth of crop and output. In the estimation of agricultural production output, the regression model is a useful technique that can predict the effects of various factors. The estimated values of the parameters in a regression model can be used for analysis and forecast of the performance of agricultural production (Jaikla et al., 2008; Merdun et al., 2006; Schaap and Leij, 1998). Regression Model has been used for estimating the behavior or performance of such crops as rice (Jaikla et al., 2008), corn and soybean (Ainong Li, 2007; Prasad et al., 2006) or wheat (Dadhwal and Sridhar, 1997).

Mathematical models, either the crop model or the regression model are statistical models, which require prior knowledge about the relationship among variables. This is in order to define the relationships between factors and effects of the data including the information or conditions which underly the model and which are being tested, proven, or examined to be true. Thus, the results are considered correct and accepted (Ainong Li, 2007).

Nevertheless, there are models, which do not need assumption or statistical conditions nor variables's relationship in advance. These include the Artificial Neural Network (ANN) and the Beyesian Belief Network (BBN).

2.3.3 Artificial Neural Network (ANN)

This is a model from which one can learn from experience or past information. The conceptual structure of this model has three processes as follows; input, output and latent (or hidden). Latent is between the input and output and can be one more. The important advantage of an ANN is that this model does not need to make any assumption about the functional relationship between variables, factors, and products. On the contrary, this model is able to learn and create a model, which is a nonlinear model itself from the relationship between factors and products with learning process. Therefore, specialists are not necessary for this method and there is not any trouble in the model process (Boonprasom and Bumroongitt, 2005). Thus, several studies concerning yield estimation have implemented the advantages within this model as an tools (Correa et al., 2009; Green et al., 2007; Kaul et al., 2005; Merdun et al., 2006).

However, the disadvantage of ANN is in the difficulty of understanding every part of the model because the model process is hidden. (called Black Box) (Correa et al., 2009). The structure that has much more hidden space will cause false data, which affects results and mistaken forecasts. (Kaul et al., 2005). Consequently, there is another method to improve and eliminate this disadvantage. It is known as the Beyesian Belief Network.

2.3.4. Beyesian Belief Network (BBN)

This is one of the most interesting ways to the development of a decision support system dealing with risks or variations by gathering data for analysis from the knowledge of specialists based on prior measured data. Moreover, this system can also take into account certain random variables to find out the possibility distribution of a variant of incidents. Then the risks and unpredictable incidents will be shown and used to support one's decisions. BBN is a tool, which can help communicate and understand the management system and the farmer's behavior during production. (Aalders, 2008, Smitha et al., 2009). BBN method is a transparent process. It can be observed from every part of the model so the user can understand and have more confidence in agreeing with the results. Furthermore, BBN can inform us of the qualitative reasons of each part of the model which can be explained by the existing data to help understand the system (Correa et al., 2009). Furthermore, BBN can be applied in the study of plant production estimation (Gu, 1994; Gu et al., 1996; Martnez-Rodrguez et al., 2008; Tari, 1996).

It is obvious that BBN is very useful to basically address the underlying situations where the data analysis is based on inconsistency conditions of both the input information and the results that are gathered.

2.4 Longan Yield Model by BBN

BBN is a cross between quantitative information and qualitative information. It is meaningful since longan yield estimates nowadays cannot be done using the Crop Model like the case of field crop or any annual plants because of the specifications of the international crop development model (Hester and Cacho, 2003). Moreover, there are other situations that affect longan production leading up to the harvest with the start from the recovery from the previous harvest. From the blooming of flowers and fruit production to plant practice or management, each step carries a high uncertainty. The information is needed about the land property, the age of the plants, sustainable knowledge of management and soil and water resources, all which can affect longan production. Therefore, the integration of longan yield estimates using the BBN, GIS, and plant knowledge along with the co-ordination of information collected by specialists and experienced farmers, and the information from fieldwork surveys can help develop a system for more efficient longan output estimation in the future.

BBN Model can spatially analyze the balance of longan management by analyzing the probability of the causes and results which occur during each step of longan production, from planting until harvesting by using lines to show the relationship between the causes and results in each situation (Bromley et al., 2005). The results from the estimates by that model can detail the causes and results of each element in the system. They also inform us the probabilities of the results from other related factors when the situation changes according to the concept and the structure of BBN estimation as shown in Figure 3.

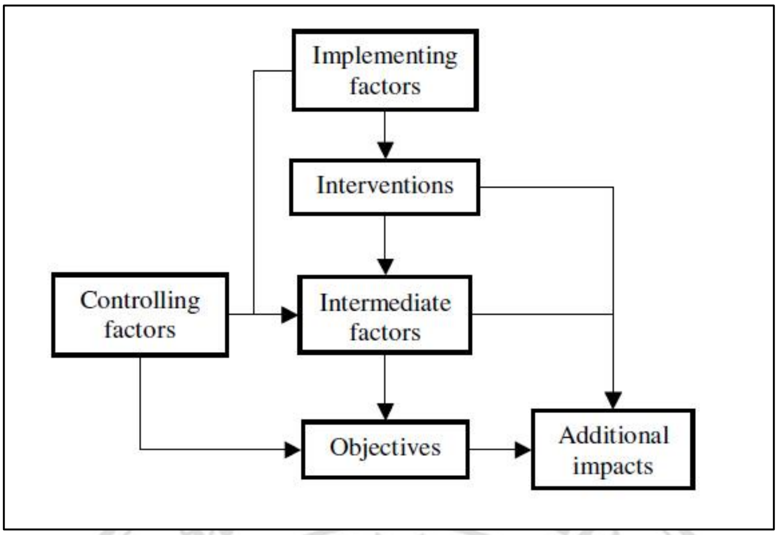


Figure 3 The structure of the network built by BBN (Cain, 2001).

The factors in the system implemented can be the cause of the occurrence of factors in other levels. The level of occurrence of a factor will depend on the type of relation in each factor. Concerning the opportunity of result occurred from cause, there is the concept of Bayes' Rule for using probability to define the relationship of each factor (Cain et al., 2003). Bayes' theorem can be written as the following equation;

$$P(B|A) = \frac{P(A|B).P(B)}{P(A)} \quad \text{………….………….. (1)}$$

P (A) is the probability that situation A will occur.

P (B) is the probability that situation B will occur.

P (B/A) is the probability that situation B will occur after situation A.

P (A/B) is the probability that situation A will occur after situation B.

And the model of the relationship of factors according to Bayes' theorem is as shown in Figure 4.

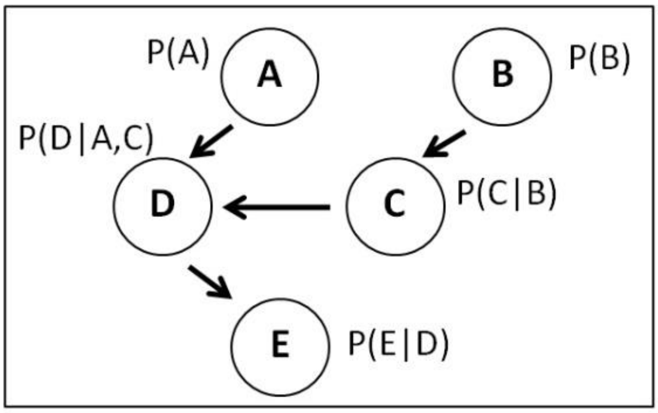


Figure 4 The probability theory used in Bayes's Network (applied from Cain, 2001).

Therefore, when the principle of creating a structural BBN network from Bayes' theorem was used with knowledge about in longan production, each implement and the direction of output in each part of the system can be shown in a graphic model, which was easily communicated and understood. It also has the capability of improvement or changes in value of opportunity in probability along all paths of the network. When it is modified for use with longan production system, it can help us understand all of the elements of the system correctly and this assists us to gain an overview of the solution.

BBN structure consists of three main components (Cain et al., 2003): (1) the set of nodes, (2) the set of links, and (3) the set of probabilities. The set of nodes represents the main parameters of the target system. Each parameter has a certain state which can be distinguished from each other clearly. Such examples of these variables or nodes are the physical, biological, socio-economic with a status of good, moderate, or poor. The second component is the set of links representing a causal relationship, herewith referred as the parent node and the effect between nodes referred as child node. The last component is the set of probabilities which indicates the value of status in each node. This value depends on the probability of the presence of adjacent source nodes. The relationship pattern of all three components is shown in Figure 5.

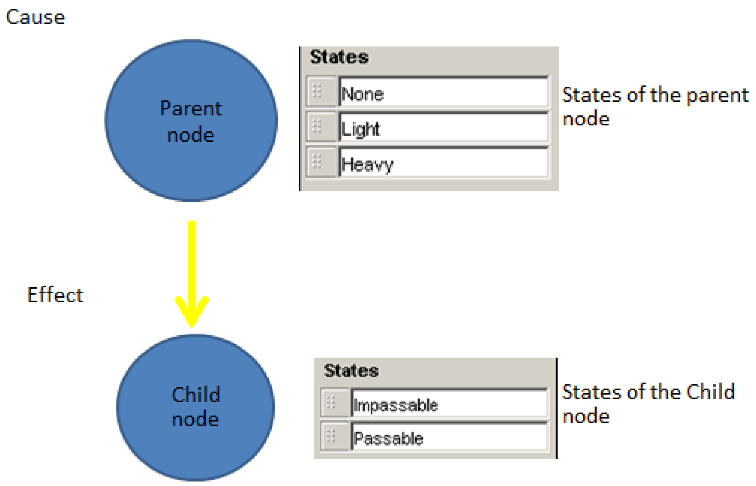
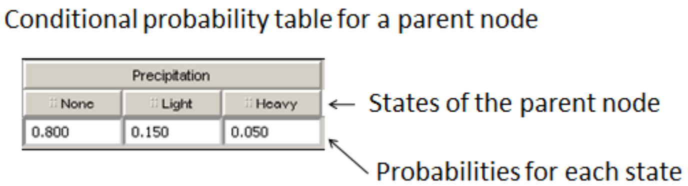


Figure 5 Link between Cause (Parent node) and Effect (Child node) (Cain, 2001).

BBN modeling is constructed from the first component and the second component in terms of relationship diagram whereas the third component is the property of the node stored in a table of probabilities (Conditional Probability Table, CPT). The original node and destination node have to define the CPT value in every state in which CPT is a level of a chance of each state in one node. An example is shown in Figure 6, where the original node has three states defined as none, light, and heavy whereas the chance of each state is 0.8, 0.15, and 0.05, respectively. The total of these three states is 1.



Precipitation		
None	Light	Heavy
0.800	0.150	0.050

Figure 6 CPT of Parent node (Cain, 2001).

When the original node is associated with the destination node, the CPT of each state in the original node is needed to calculate with the CPT of the destination state. The model of this relationship was already defined. The result of calculation of CPT in both nodes is shown in Figure 7.

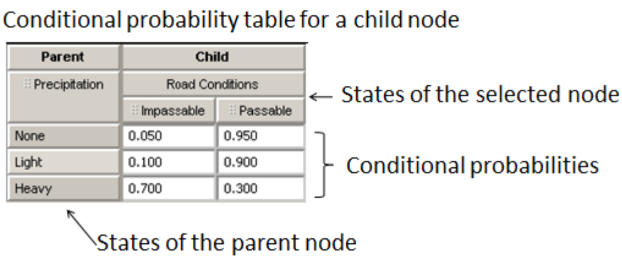


Figure 7 CPT calculation of Parent node and Child node (Cain, 2001).

A relationship between two nodes in Figure 7 is explained by that the state in the original node called “none” has effect on the chance of state in destination node and there are two states: impassable and passable. The value of impassable equal to 0.05, whereas passable equal to 0.95. It can be seen that the configuration of the connection of each factor in the model is based on the calculation of the total CPT (David et al., 1993, Bouissou et sl., 1999).

The CPT value may have been created from relationship obtained from academic reports; i.e. by asking the expert or experienced farmers in the target production. After a CPT has been created completely as well as the BBN Model Test, that model is ready to be used for analysis. However, the user might try to change the probability of any of the status levels of the nodes. The model will calculate the Joint Conditional Probability (JCP) value of the nodes in the BBN model according to Bayes’ Theorem (Varis, 1997).

The BBN Model can be created with a packaged program such as the Netica software (www.norsys.com), which was developed for node management and flow chart relationships. It is convenient to create the CPT, calculate the JCP, and to conduct sensitivity analysis. Netica program which typically contains a graphical user interface (GUI) for modeling and inference engine based on the junction tree algorithm for computation (Norsys Software Corp, 2011).

However, in order to achieve a more definite estimate of longan yield in advance, this research has added an important implement, that is to determine the quantity of fruit in the canopy at least one month before harvest. Thus, this study has devised a shortcut method to find out how to count the fruits in the sampled trees and to use that data as an input in the BBN Model.

2.5 The Measurement of Longan Fruit in the Bush by Digital Photo

There are several pieces of research that use digital photo as main input data along with the data recordings of several other devices for example Bulanon et al.'s Research (2009) used fusion imaging, which records a thermal image with an infrared camera and a visible image taken with a digital camera in Red Green Blue mode. Both images were fused to distinguish the orange fruits from the leaves in the canopy. The result revealed that fused image focused on the orange fruits more clearly than the leaves (Bulanon et al., 2009). Additionally, there were several pieces of research that used digital photo for analysis as an input factor to approach the process of fruit yield estimation in advance like Stajnko et al. (2004). This study was conducted to find a method of estimating the fruit yield in advance by using an image in RGB mode to create a Normalized Difference Index (NDI) to make more differences between value of leaves and fruits. Therefore, they could count the number of apples and calculate the size better than using human laborers.

Okamoto et al. (2009) studied how to develop the tool in the process of image analysis to assess the number of oranges in each tree before the harvest or when the oranges are not mature and still have a green peel. They used this information to plan for the harvest in advance by using color images, which were taken with a hyper-spectral camera. It was able to record in the visible bandwidth and the low-infrared bandwidth (the length of wave being about 700-1,000 nm). Later, the digital images were brought for analysis, in two processes: pixel spectral processing and spatial processing. The pixel spectral processing was to do pixel segmentation and separate the unrelated background. Spatial processing was then used for fruit object detection. It was used to separate the oranges from leaves in the canopy. The results that were found had a high level of accuracy but it was not good enough for output estimation. The cause of error was that too many young green leaves showed up in the picture. Therefore, it was difficult to separate green oranges from green leaves. Finally, they used the non-linear discriminate method in this analysis and they got more accurate results (Okamoto and Lee, 2009).

From all the researches reviewed above, these were just some studies that describe the ability of digital imaging taken with normal cameras and used for analysis and separation of the composition. These processes, the means of image separation, and the way they are used with other recording devices for the analysis of fruit yield estimation gave the best, the most accurate and most satisfying results. They could be used as a guideline for this study as well.

The combination of all the tools mentioned above can be used to estimate a crop yield. Typically, the error of instruments can be checked by using a root mean square error (RMSE) which measures the differences between values predicted by a model or an estimator and the values actually observed. The RMSE in crop yield estimation model less than 10% is considered as very high predictive accuracy, whereas the value between 10% - 20% considered as high accuracy and between 20 - 30% considered moderate accuracy. A model will not be robust for the lack of accuracy if its RSME is high than 30% (Timsina et al, 2006; Buddhaboon, 2011).