Searching for simplified farmers’ crop choice models for integrated watershed management in Thailand: A data mining approach

Benchaphun Ekasingh a,b,*, Kamol Ngamsomsuke a

a Department of Agricultural Economics, Faculty of Agriculture, Chiang Mai University, Thailand
b Multiple Cropping Center, Faculty of Agriculture, Chiang Mai University, Thailand

Abstract

This study used the C4.5 data mining algorithm to model farmers’ crop choice in two watersheds in Thailand. Previous attempts in the Integrated Water Resource Assessment and Management Project to model farmers’ crop choice produced large sets of decision rules. In order to produce simplified models of farmers’ crop choice, data mining operations were applied for each soil series in the study areas. The resulting decision trees were much smaller in size. Land type, water availability, tenure, capital, labor availability as well as non-farm and livestock income were found to be important considerations in farmers’ decision models. Profitability was also found important although it was represented in approximate ranges. Unlike the general wisdom on farmers’ crop choice, these decision trees came with threshold values and sequential order of the important variables. The decision trees were validated using the remaining unused set of data, and their accuracy in predicting farmers’ decisions was around 84%. Because of their simple structure, the decision trees produced in this study could be useful to analysts of water resource management as they can be integrated with biophysical models for sustainable watershed management.

1. Introduction

Land, water and forest resources in Thailand are becoming scarce and depleted. Due to increased population growth and pressure for farmlands, large proportions of the watersheds have been converted to agricultural use. In many of the Thai watersheds, much of this conversion has occurred on high slopes. Soil erosion and soil fertility have become major problems for not only farmers, but also the wider community due to their effects on water quality. An issue related to the above-mentioned problems is how upland people use and manage their land. Proper land utilization and management of watersheds is important for the whole society. Water resources have been substantially reduced both in quantity and quality. Watershed management that takes into account the needs of uplanders and lowlanders requires an appropriate integrative approach and tools.

The Integrated Water Resource Assessment and Management (IWRAM) Project aimed to create an integrative decision support system (DSS), through the integration of agronomic, hydrological, soil erosion and socio-economic models (Jakeman and Letcher, 2003; Jakeman et al., 2005). The IWRAM project started in 1997 with collaboration of scientists from Australian National University and Thai researchers from Ministry of Agriculture and Agricultural Cooperatives and other Universities. The project has involved three phases, each with its own study areas within Thailand. Phase I covered a watershed called Mae Chaem in Chiang Mai Province, Northern Thailand during 1997–2000 (Letcher et al., 2006a, 2006b; Merritt et al., 2004). Phase II covered the Upper Ping II watershed in Chiang Mai and Lamphun provinces during 2001–2004. Phase III covered two larger Mekong watershed areas — the Chan Watershed in Chiang Rai province and the San Watershed in Loei province during 2004–2006. While the conceptual framework of the project (Fig. 1) stayed the same, the details of each phase differed as the integrated components were continuously improved.

Hydrological models incorporating certain land uses were developed to predict water yield which would be an input into a crop model to predict crop yields. Based on chosen crops and yields, soil erosion estimates could be obtained from a soil erosion model. A set of key indicators of sustainable watershed management were monitored. A crucial link to biophysical models is on
farmers’ crop choice which determines land use. Letcher et al. (2005) outlined some approaches to socio-economic impacts and decision processes for the project.

Some farmers’ crop choice models using data mining were produced (Ekasingh et al., 2005; Letcher et al., 2005; Ngamsomsuke and Ekasingh, 2007). Nevertheless, the decision trees produced in the study were found to be somewhat complex and large in size. For example, in Ekasingh et al. (2005), the farmers’ decision tree (DT) for wet season cropping in the Upper Ping watershed consisted of 29 leaves and the same number of leaves resulted in the DTs for the Chan and San Watershed areas (Ngamsomsuke and Ekasingh, 2007). For the dry season, the studies resulted in farmers’ DTs that consisted of 14 leaves in Upper Ping and 21 in the Chan and San Watersheds. In the 2007 study, the farmers’ DTs were developed from a combination of the data from the two watersheds. This was a major limitation to the DTs as the two watersheds differed in their cropping patterns and socio-economic attributes. Also, these DTs made use of many dichotomous groupings of data e.g. many were classified as “equal to” vs “not equal to” and in many cases three or more groups were probably more appropriate for classification. Nevertheless, despite these weaknesses, the resulting DTs were able to achieve high correct classification rates (around 87% or higher). These accuracy rates are test statistics produced by WEKA data mining program by using the DTs to predict outcomes from unused datasets. Feedback from and consultation with biophysical scientists and policy makers highlighted the need for better understanding and simplification of farmers’ crop choice for further integration in the current watershed decision support system.

The objective of this study is to simplify models of farmers’ crop choice previously constructed by our team so that its relevance to integrated watershed management and policies can be improved.

2. Data mining techniques and knowledge discovery

Knowledge discovery through data mining, machine learning, artificial neural networks or related methods is becoming common practice in many fields e.g. including business (e.g. Shaw et al., 2001; Hui and Jha, 2000), health (e.g. Kunene and Weistroffer, 2008; Alonso et al., 2002; Belanche-Muñoz and Blanch, 2008), agriculture (e.g. Waheed et al., 2006; Delgado et al., 2009) and natural resource management (e.g. Kalteh et al., 2008; May et al., 2008). Like its related methods, data mining is an iterative learning process that involves discovering previously unknown and potentially interesting patterns in large datasets (Piatetsky-Shapiro and Frawley, 1991). The selection of data mining algorithms, hypotheses formation, model evaluation and refinement are key components of this discovery process. The mined information is represented as
Data mining procedures and techniques have been discussed elsewhere e.g. in Buntine (1993), Quinlan (1993, 1996) and Witten and Frank (2005). The procedures are essentially machine learning techniques for inducing domain models or analyzing datasets. Decision trees are an output of the procedures and there are many algorithms available for constructing decision trees. The algorithm used in this paper is the C4.5 model (Quinlan, 1993). C4.5 is a commonly used standard algorithm to produce decision tree models. The decision tree models produce ordered sets of rules. The C4.5 classification scheme is a univariate decision tree approach that uses a divide and conquer method to split the dataset, based on the values of a feature that is the best choice according to a selection measure (Gain Ratio) (Quinlan, 1993, 1996; Windeatt and Ardehali, 2004).

The C4.5 algorithm is available in the J48 classifier tree in the WEKA (Waikato Environment for Knowledge Analysis) software package Version 3.5.8 developed by Waikato University, New Zealand (Witten and Frank, 2005; University of Waikato, 2008). The WEKA software is well known in the data mining community and is
available free of charge from the world wide web. Through WEKA, Cunningham and Holmes (1999) indicated that information “mined” from data can provide insights into the domain being studied that run counter to the received wisdom of a field. Locating these surprising or unusual portions of the model can be the focus for a data mining analysis. In some data mining applications, the goal might be to use a model as a predictive tool, to provide automated classification of new instances.

3. The study site

The study was conducted using field data in two watersheds: the Chan Watershed in Chiang Rai province (Northern Thailand) and the San Watershed in Loei province (Northeastern Thailand). The Chan and San Watersheds drain into the Mekong River. The Chan Watershed covers part of Dan Sai and Phu Roer districts, Loei province. The watershed is mostly undulated mountainous areas. There are some small flat areas located between the high mountainous slope areas. Many of its streams drain into Mekong River. The elevation of the watershed ranges from 320 to 1365 msl. The watershed covers 861 km$^2$ and like the Chan Watershed is extensively cultivated by farmers. 54% of the watershed is used for farming, 40% is forested, 3.5% are grass areas and 2.5% are residential areas. Compared to the Chan Watershed, there are a limited number of crops grown in the areas. Maize is a popular crop. Upland rice, glutinous rice, non-glutinous rice, ginger, cassava, lychee, tangerine and sweet tamarind are also grown. The San Watershed, unlike the Chan Watershed, is not densely populated, and farm sizes are generally larger. There were about 72,750 people living in the San Watershed in 2003.

The Royal Forest Department (RFD) has a duty to protect both the Chan and San Watersheds from further encroachment because strictly, most of the watershed areas are legally defined as forest areas. Nevertheless, it is a traditional practice for farmers to clear land for agricultural purposes especially when population pressure is increasing. Some compromises have been negotiated between RFD and communities as to how much of the watershed areas can be used for farming. However, both watersheds have experienced problems of forest encroachment continuing beyond that agreed.

In Phase III of the IWRAM project, the research team decided that land classification would be based on the soil series as defined by the Thai Department of Land Development (DLD) rather than on resource management units (RMU) or land management units (LMU) defined by the IWRAM research team in Phase I or II respectively. This was because the DLD soil series names and among highland farmers. Dry season cropping is possible in areas where there is water availability.

San Watershed covers part of Dan Sai and Phu Roer districts, Loei province. The watershed is mostly undulated mountainous areas. There are some small flat areas located between the high mountainous slope areas. Many of its streams drain into Mekong River. The elevation of the watershed ranges from 320 to 1365 msl with the average of 842.5 msl. The watershed covers 861 km$^2$ and like the Chan Watershed is extensively cultivated by farmers. 54% of the watershed is used for farming, 40% is forested, 3.5% are grass areas and 2.5% are residential areas. Compared to the Chan Watershed, there are a limited number of crops grown in the areas. Maize is a popular crop. Upland rice, glutinous rice, non-glutinous rice, ginger, cassava, lychee, tangerine and sweet tamarind are also grown. The San Watershed, unlike the Chan Watershed, is not densely populated, and farm sizes are generally larger. There were about 72,750 people living in the San Watershed in 2003.

The Royal Forest Department (RFD) has a duty to protect both the Chan and San Watersheds from further encroachment because strictly, most of the watershed areas are legally defined as forest areas. Nevertheless, it is a traditional practice for farmers to clear land for agricultural purposes especially when population pressure is increasing. Some compromises have been negotiated between RFD and communities as to how much of the watershed areas can be used for farming. However, both watersheds have experienced problems of forest encroachment continuing beyond that agreed.

In Phase III of the IWRAM project, the research team decided that land classification would be based on the soil series as defined by the Thai Department of Land Development (DLD) rather than on resource management units (RMU) or land management units (LMU) defined by the IWRAM research team in Phase I or II respectively. This was because the DLD soil series names and

### Table 2

Results of data mining.

<table>
<thead>
<tr>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location</td>
<td>San Watershed</td>
<td>Chan Watershed</td>
</tr>
<tr>
<td>Soil series</td>
<td>Ky</td>
<td>AL</td>
</tr>
<tr>
<td>Season</td>
<td>Wet</td>
<td>Wet</td>
</tr>
<tr>
<td>Instances</td>
<td>170</td>
<td>212</td>
</tr>
<tr>
<td>Variables</td>
<td>12</td>
<td>15</td>
</tr>
<tr>
<td>No. of leaves</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>Size of trees</td>
<td>21</td>
<td>21</td>
</tr>
<tr>
<td>Variables tested</td>
<td>LandType, Tenure, Water, Flood, Drought, CashCost, CropType, Labor, LandLabor, Livestocks, NonFarmIncome, Loan</td>
<td>AreaPlanted, LandType, Tenure, Water, Flood, Drought, CashCost, E_GMargin, HHmem, Labor, FarmSize, LandLabor, Livestocks, NonFarmIncome, Loan</td>
</tr>
<tr>
<td>Correctly classified instances</td>
<td>84.10%</td>
<td>83.90%</td>
</tr>
<tr>
<td>High accuracy</td>
<td>Lychee, sweet tamarind, upland rice, non-glutinous rice and ginger (100%)</td>
<td>Maize, sweet corn, orange (100%), non-glutinous rice (92%)</td>
</tr>
<tr>
<td>Medium accuracy</td>
<td>Lychee (84%), glutinous rice (67%)</td>
<td>–</td>
</tr>
<tr>
<td>Low accuracy</td>
<td>Glutinous rice (50%) and maize (42%)</td>
<td>–</td>
</tr>
</tbody>
</table>

### Table 3

Confusion matrix for wet season, San Watershed.

<table>
<thead>
<tr>
<th>Actual</th>
<th>Predicted</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Instances classified as</td>
</tr>
<tr>
<td></td>
<td>Lychee</td>
</tr>
<tr>
<td>Lychee</td>
<td>26</td>
</tr>
<tr>
<td>Sweet tamarind</td>
<td>0</td>
</tr>
<tr>
<td>Upland rice</td>
<td>0</td>
</tr>
<tr>
<td>Non-glutinous rice</td>
<td>0</td>
</tr>
<tr>
<td>Maize</td>
<td>1</td>
</tr>
<tr>
<td>Ginger</td>
<td>0</td>
</tr>
<tr>
<td>Glutinous rice</td>
<td>0</td>
</tr>
</tbody>
</table>

### Table 4

Confusion matrix for Chan Watershed, wet season.

<table>
<thead>
<tr>
<th>Actual</th>
<th>Predicted</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Instances classified as</td>
</tr>
<tr>
<td></td>
<td>Lychee</td>
</tr>
<tr>
<td>Lychee</td>
<td>27</td>
</tr>
<tr>
<td>Maize</td>
<td>0</td>
</tr>
<tr>
<td>Sweet corn</td>
<td>0</td>
</tr>
<tr>
<td>Orange</td>
<td>0</td>
</tr>
<tr>
<td>Glutinous rice</td>
<td>2</td>
</tr>
<tr>
<td>Non-glutinous rice</td>
<td>0</td>
</tr>
</tbody>
</table>
definitions are standardized and nationally recognized in Thailand. This will ease model extension nation-wide.

According to the DLD soil classification, there are eleven soil series in the Chan Watershed: Alluvial (AL), Ban Chong (Bg), Chiang Rai (Cr), Chiang Saen (Ce), Hang Dong (Hd), Mae Sai (Ms), Nong Mot (Nm), Phimai (Pm), Tha Muang (Tm), Thad Panom (Tp) and Slope complex (SC). There are eight soil series in San Watershed: Dan Sai (Ds), Hang Dong (Hd), Khao Yai (Ky), Loei (Lo), Lat Ya (Ly), Ngao (No), Ta Yang (Ty) and Slope complex (SC).

4. Data collection

In Phase III of the IWRAM project, a survey of 340 and 167 farm households from the Chan and San Watersheds was conducted in 2005 based on DLD soil series map. As the project involved hydrological and soil erosion models, exact coordinates of data points were considered necessary. Global Positioning System equipment together with detailed administrative maps were employed to pinpoint the exact location of plots to be surveyed, based on soil series. Farmers who owned those plots were sought and interviewed. This process proved demanding in terms of time and costs of field survey but had enabled the datasets to be spatially linked. Structured questionnaires covered such variables as household characteristics, land tenure and utilization, crops grown in different seasons, problems of farming, plot characteristics, production cost and revenue, financial support, use and management of irrigation systems and environmental problems, etc. Data mining analysis was conducted on this dataset.

5. Data mining procedures used

In this study, data mining procedures were applied to the dataset of each of the two watersheds to produce decision rules for farmers’ wet and dry season crop choice. Fig. 2 shows the factors influencing farmers’ decisions in crop choice as found by others (Thong-ngam et al., 1997; Ekasingh et al., 2005; Shinawatra et al., 1987; Shinawatra, 1988). The variables used as possible classifiers of crop choice model for the data mining analysis in this study are presented in Table 1. These variables were based on those found by Ekasingh et al. (2005) and Ngamsomsuke and Ekasingh (2007), although some groupings were revised. The data mining algorithm was applied to determine which variables were significant at what threshold levels and the order in which they appeared. The datasets were prepared for each soil series. This paper presents the results of data mining for the main soil series of the study areas e.g. Ky soil series in the San Watershed and AL soil series in the Chan Watershed.

6. Farmers’ crop choice decision trees

Osei-Bryson (2004, 2007) suggested if a DT is to be used as both an explanatory and predictive model, it is important that the DT should be as simple as possible. Tree simplicity is therefore emphasized here and is represented by number of leaves as well as tree size. The percentage of correctly classified instances will be lower showing a tradeoff for smaller trees.

The resulting DTs (see Figs. 3–5) were much smaller in size and simpler than earlier DTs presented in Ekasingh et al. (2005) and Ngamsomsuke and Ekasingh (2007). In the 2007 study, the authors used the same dataset but the number of leaves was 112 in the wet season and 23 for the dry season. Even when the minimum number

---

Table 5
Confusion matrix for Chan Watershed, dry season.

<table>
<thead>
<tr>
<th>Actual</th>
<th>Predict</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Non-glutinous rice</td>
</tr>
<tr>
<td>Non-glutinous rice</td>
<td>5</td>
</tr>
<tr>
<td>Sweet corn</td>
<td>0</td>
</tr>
<tr>
<td>Soybean</td>
<td>0</td>
</tr>
<tr>
<td>Maize</td>
<td>0</td>
</tr>
<tr>
<td>Glutinous rice</td>
<td>0</td>
</tr>
<tr>
<td>Tobacco</td>
<td>0</td>
</tr>
</tbody>
</table>

---

Fig. 6. Farmers’ typology and crop choice, San Watershed, wet season.
of objects per leaf was increased (35 for the wet season and 5 for the dry season). Ngamsomsuke and Ekasingh (2007) still ended up with 29 leaves for the wet season and 21 for the dry season. The 2005 studies also had large DTs. Through an iterative process the final DTs in this study were reduced in size having 12 leaves in the wet season and only 7 in the dry season (see Table 2). Using the 10-fold cross-validation method in WEKA, the percentage of correctly classified instances was around 84%. Land types, land tenure, cash cost, and access to loan were found to be important variables in wet season models. In the wet season models, irrigation facilities, livestock income and land–labor ratio were important in the San Watershed while in the Chan Watershed, flood problem, expected gross margin, size of household members and labor were classifier variables. In the dry season, non-farm income, drought problem, access to loan and land–labor ratio were classifier variables.

Using the resulting DTs, high levels of accuracy of prediction of instances (92–100%) were obtained for cash crops like lychee, sweet tamarind, ginger, non-glutinous rice in the San Watershed areas in the wet season; orange, upland rice, non-glutinous rice, maize and sweet corn in the Chan Watershed in the wet season; and glutinous rice, soybean and tobacco in Chan Watershed in the dry season. However the DTs did not accurately predict the choice of glutinous rice and maize in the wet season in the San Watershed and non-glutinous rice in the Chan Watershed in the dry season (42–50% level of accuracy) (Tables 2–5). Each DT revealed useful information on farmers’ crop choice and their use of watershed areas. The data mining exercises proved to be valuable in enhancing understanding of how and why farmers use their farmlands in particular ways. The following are the detailed description of the DTs.

---

**Fig. 7. Farmers’ typology and crop choice, Chan Watershed, wet season.**

**Fig. 8. Farmers’ typology and crop choice, Chan Watershed, dry season, irrigated areas.**
Model 1. San Watershed, wet season

If farmers would like to grow a high value crop (which is more risky), they would grow lychee. If not, they can choose to grow lychee, rice, sweet tamarind and maize.

To choose among the latter crops,

a) if farmers have less than 4.67 rai per person and have land ownership, they would grow lychee in their land, otherwise they will grow upland rice;
b) if farmers have more than 4.67 rai per person, they can choose to grow lychee, non-glutinous rice, glutinous rice, sweet tamarind or maize with the following conditions:

b.1) if they have income from livestock, they will choose to grow non-glutinous rice, otherwise they can choose among glutinous rice, maize and sweet tamarind.
b.1.1) Maize needs more cash investment than rice or sweet tamarind. If farmers have the required capital (more than 2000 baht per rai), farmers will choose maize. For crops with a cash investment less than 2000 baht, farmers can choose glutinous rice or sweet tamarind; farmers may choose maize if they can borrow money.
b.1.2) In the paddy land, farmers will always grow glutinous rice. In the lowland, if farmers have flood problems, they would grow sweet tamarind.
b.1.3) In the upland, farmers will grow sweet tamarind. If farmers have large land (48 rai per person), they will rather grow sweet tamarind, otherwise they will grow maize although with lower levels of investment.

Model 2. Chan Watershed, wet season

If farmers are willing to pay for higher levels of cash investment in crop production (more than 4000 baht per rai), they would choose to grow orange, otherwise they can grow either rice, lychee, sweet corn or maize.

To choose among the latter crops,

a) If they are looking for an expected gross margin more than 5000 baht per rai, they should grow non-glutinous rice. Non-glutinous rice nevertheless requires more capital, and it is also not usually preferred by Northern farmers unless they already have some glutinous rice in stock for home consumption.
b) If they are not necessarily looking for more than 5000 baht of gross margin, they can use the following decision rules,
b.1) in paddy land, they would grow glutinous rice, if they have little labor and capital. If they have more capital, they can grow non-glutinous rice. If they have a lot of labor, they will grow glutinous rice;
b.2) in the lowland, if there is no flood in their land, they will grow lychee. If they experience some flood and have a larger family size, they will grow lychee. If they have a small family, they will grow sweet corn;
b.3) in the upland, if they have ownership of the land and no flood, they will grow lychee. But if they have ownership of the land but experience occasional flooding on their land, they will grow sweet corn. If they do not have ownership of their land, they will grow maize;
b.4) on the sloping land, they will grow lychee.

Model 3. Chan Watershed, dry season

In paddy land, if farmers do not have non-farm income (no cash to buy rice), they will grow glutinous rice. If they have non-farm income and do not have drought problems, they would grow non-glutinous rice but if they have problems with drought, they would grow soybean.

In the lowland, if farmers are able to loan more than 50% of total investment required, they would grow tobacco. If they cannot loan more than 50% of total investment requirement, they would grow either sweet corn or maize. They will grow maize if they have more land per labor (>2.88 rai per person) and will grow sweet corn if they have less land.

In the upland, they will grow maize.

7. Model applications

These DTs revealed surprisingly detailed information about farmers’ crop choice which was previously unknown. They showed interaction between biophysical conditions such as floods and drought with crop choice decisions. Moreover, they revealed what socio-economic characteristics e.g. land ownership, access to loan, attitude to risk (via expected income), off-farm income, etc., influence crop choice decisions and at what degree of influence. Previous knowledge of farmers’ crop choice was not able to provide such detailed patterns of how farmers make their choice of crops. The resulting DTs were not only results of machine learning algorithms, but rather results of a combination of machine learning (data mining) and expert inputs.

From the resulting DTs, many policy implications of integrated watershed management can be drawn. For example, in the San Watershed, the DT highlights the importance of land tenure, without which farmers are not willing to invest in high-investment lychee orchards. Land tenure policy in watershed areas will need to be revisited to enable farmers’ security of their farmlands. The DT also shows that access to a threshold level of loan determines preference to such cash crops as maize over the low-investment-low-return sweet tamarind. Here, the agricultural loan policy on the maximum loan size can be potentially examined and revised. The DT also reveals that the existence of other income earning opportunities, e.g. livestock income, enables commercial farmers to grow non-glutinous rice for sale, rather than their traditional glutinous rice for consumption. Policy on promotion of livestock and off-farm employment can be refined.

Following a farming systems method called farmers’ typology as exemplified by Thong-ngam et al. (1997), Figs. 6–8 reinterpret the DTs with further analysis on farmers’ typology (subsistence, semi-commercial and commercial farmers). Analysis of farmers’ types and their resources will be necessary to predict farmers’ crop choice in each type of soil. Fig. 5 shows an analysis of farmers’ typology based on the DT in San Watershed. Figs. 7 and 8 display similar analysis for Chan Watershed in the wet and dry season. In the Chan Watershed, like in the San Watershed, land tenure is an important consideration in
farmers’ decisions on crop choice. Farmers would not grow fruit trees like lychee in the uplands unless they have ownership to their land; annual crops such as maize are grown instead. Lychee is planted in the uplands or lowlands preferably in areas with no flooding problems. Good drainage is required to grow lychee. To grow other fruit trees e.g. orange, farmers need to have good levels of capital as this requires high investment. Commercialized farmers with good access to capital would grow non-glutinous rice while subsistence farmers with less access to capital and lower preference to borrow capital would normally grow glutinous rice.

In the dry season in the Chan Watershed, good access to capital will enable farmers to grow non-glutinous rice. This reiterates the importance of agricultural credit policy. A crop like tobacco is a contract farming crop and is mostly adopted by commercial farmers in the irrigated lowlands with the support of credit from tobacco companies. Non-farm income, like livestock income, will provide capital to farmers for the planting of soybean, non-glutinous rice and tobacco and also provide farmers with income to buy glutinous rice for consumption. For semi-commercialized farmers, without non-farm income, farmers need to grow glutinous rice as they would have no cash to buy rice for consumption.

The DT for the dry season can inform policy makers of the importance of water scarcity in the paddy. Water scarcity compels farmers to grow soybean instead of non-glutinous rice.

In the wet season DTs for both watersheds, crop profitability (e.g. gross margin) is also an important consideration for farmers’ adoption especially when farmers are commercialized or semi-commercialized.

While general knowledge on what factors influencing farmers’ crop choice is not lacking, the data mining approach and the resulting DTs provide threshold values as well as the order of important variables in the decision process. Given such knowledge discovery, scenarios can be constructed with respect to the policy environment in order to project the trends of land use and the resulting impact on the watersheds under investigation. Scenarios on particular policies e.g. on land tenure, credit, watershed zoning, etc., can be tested and their impact assessed.

8. Conclusion

This study presents simplified DTs from data mining procedures and the DTs will be much more understandable to policy makers as well as agricultural and natural resource scientists than the previously produced DTs of farmers’ crop choice. The models shed light on patterns of farmers’ crop choice. They can be linked to biophysical models for integrated watershed decision support systems so that further simulation of watershed management can be facilitated. They give importance to key socio-economic variables as much as biophysical variables. Land type, land tenure, water availability, incidence of floods and droughts, expected gross margin, cash requirements, access to capital, land–labor ratio and other income earning opportunities are factors determining farmers’ crop choice. It is expected that watersheds are different in their crop composition and it is unlikely that DTs in one watershed can be readily used in another. Nevertheless, this procedure was found to be simple and can be repeated easily and applied to other study areas.

Acknowledgement

The authors would like to appreciate all researchers and personnel in the IWRAM Project (Thai and Australian partners), Agricultural Science Foundation and the Royal Project Foundation for the research support. Thanks to Thailand National Research Council which provided the research grants for the third phase of the IWRAM project.

References


