"粒粒" Farm Crop Optimiser

1. Introduction

1.1 Farming Landscape

In China, a large majority of farms are characterised by small land plots of less than half a hectare, and worked by farmers of low education and literacy levels. In a landscape of population explosion and burgeoning demand for crop produce, the Chinese government has recognised the need to reform the traditionally inefficient agriculture in China.

1.2 Farm Crop Optimiser – Lili (**“粒粒”**[1])

About 60km North of Hangzhou, nestled in a small town near the Mogan Mountain, is a typical family-sized Chinese farm called 简法农庄 (Appendix A – Background of 简法农庄). Driven by his mission to feed Chinese families with safe and nutritious vegetables, Mr Hu, the farm owner, firmly embraces ancient Chinese farming philosophies by growing only in-season, organic food and selling them via his own e-commerce platform. However, within these 6 years of operations, Mr. Hu has never had a profitable year. Given the huge variety of crops to be planted, each with a unique seasonality schedule, and limited resources such as land acreage and human labour availability, we observed that Mr. Hu’s crop planning approach has always been unsystematic and haphazard.

1.3 Formulation of Model

In this project, our group wishes to develop a crop planning model to help make farmers like Mr. Hu make crucial and optimised decisions. Given several farmer-specified conditions, Lili will find an optimal vegetable mix, as well as their corresponding planting and harvesting period, to maximise a farmer's profit.

Our Vegetable Database consists of 30 common Chinese vegetable types (Appendix B – List of Vegetable Crops). The information is mixture of reasonable guesses based on personal observation from Mr. Hu's farm and online crop information data. Each vegetable is distinguished by several key features:

<u>Yield</u>

The yield *(Ton/Mu)* is defined as yield per ton of vegetable planted in one Mu (亩). Vegetables within the same categories have similar yield output

<u>Revenue</u>

The revenue *(RMB/Ton)* is defined as the amount of RenMinBi (RMB) received per ton of vegetable sold in the market.

<u>Growth Duration and Planting Period</u>

These vegetables have seasonal planting periods (indicated by "1") and growth durations (in months). For instance, kale can only be grown in January – February and September – December. As its growth duration is 3 months, kale that is planted in January is only ready for harvest in April. According to

<u>Planting and harvest Labour</u>

The labour required to plant and harvest each vegetable is defined by labour hours per mu per month. Vegetables within the same categories have similar labour requirements.

<u>Cost</u>

The crop cost *(RMB/Mu)* is defined as the amount of Chinese RenMinBi (RMB) required per unit area.

[1] Inspired by the classical ancient Chinese poem “锄禾日当午，汗滴禾下土。谁知盘中餐，粒粒皆辛苦”

Farmer-input Data

In addition, our model can be customised to the unique characteristics of each farm. Farmers can input their farm-specific data to make the model more relevant to their needs. These include:

1. Number of farming months
2. Initial month
3. Farm size (mu)
4. Minimum crop variety
5. Labour type - regular or seasonal. Regular labour refers to a supply of labour hours (h) that is constant throughout the year. Seasonal labour refers to an irregular supply of labour hours (h) throughout the year.
6. Available labour hours per month

1.4 Key Assumptions of Model

- Vegetables must be planted in integer units of Mu (i.e. 1, 2, 3 Mu etc).
- Vegetables cannot be harvested when they are still growing (within growth duration) and there is no partial harvest i.e. vegetable planted in a unit of land will be fully harvested at the end of its growth duration.
- All harvested vegetables can be sold in the market, hence revenue is certain.
- Productivity of land and labour remains constant throughout the farming period.
- Yield of each type of vegetable in the same category is very similar (affected to the same extent by weather, disease, theft, etc).
- Planting is done on the first day of each month and harvesting is done on the last day of each month.
- Planting and harvesting can be performed instantaneously. (E.g. Assuming that crop A has a growth duration of 2 months, if it is planted on the first day of January, then it can then be harvested on the last day of February. Thereafter, crop B can be planted on the first day of March.)

2. "粒粒" Model

Our model is broken down into two sub-models. In the first part, the Stochastic Model attempts to determine the optimal budget allocation in each crop category based on the expected risk and return values. These results serve as input into the second part of the model, the Scheduling Model. Here, the unique farming constraints are also taken into account to suggest an optimal crop mix and schedule (Appendix C- User Flow of Model). The crop categories are ideally selected based on vegetables with similar trends of returns due to factors like similar seasonal demand (example of this categorization which we use later in our model can be found in Appendix B). Such categorization would be particularly useful when granular data about every single type of vegetable may not be available, and only their broad categories.

2.1 Stochastic Model

Yield for each vegetable varies each month depending on various internal and external factors such as weather, presence of pests, diseases and resource availability. Hence, Mr Hu needs to model the yield of vegetable in each of the 4 determined categories based on historical data collected in "stochastic.csv" to ascertain the categories which meets more than his minimum expected returns.

2.1.1 Decision Variables

X: a 1-dimensional array; specifies budget allocation to each crop category, Y: dual variable 1, Z: dual variable 2

2.1.2 Objective

- T is the total period of historical returns available
- N is the number of stochastic categories
- $\propto$ is the % CVaR

The objective function minimizes risk at 10% CVAR over 4 crop categories.

Min $\propto * T * z + \sum_{i=1}^{T} y_i$

2.1.3 Constraints

2.1.3.1 Budget Constraint

This indicates that total budget allocated to each category should not exceed total budget available.

$$\sum_{i=1}^{N} x_i \leq Budget, where\ x \geq 0$$

2.1.3.2 Expected Returns

This indicates that sum of the mean return of all chosen categories should be more than the expected returns.

$\sum_{i=1}^{N} RetMean_i * x_i \geq expReward$, *where RetMean$_i$ is the mean return from stochastic category i, and expReward is the expected reward.*

2.1.3.3 Dual Formulation

This helps to linearize the problem to program for linear optimisation.

$z + y_i \geq \frac{-1}{\propto * T} \sum_{j=i}^{N} RetTable_{ij} * x_j\ \forall\ i \in \{1, \ldots, T\}$, *where* $RetTable_{ij}$ *is the historical returns of each stochastic category i over month j*

$y \geq 0$

2.2 Scheduling Model

The Scheduling Model selects vegetables only from the chosen categories obtained in the Stochastic model.

2.2.1 Variables

From the "InputFile" CSV, we read the following variables for each crop: YieldPerMu, RevPerTon, GrowthDuration(months), PlantLabourHrPerMu, HarvestLabourHrPerMu, PlantingSeasons(months) and CostPerMu. (Refer to Appendix D)

2.2.2 Decision Variables

X: a 2-dimensional array; this specifies the amount of vegetable i to plant in month j

Y: a 1-dimensional array; this specifies the amount of land space available for planting in month j

Z: a 2-dimensional array; this specifies the amount of vegetable i to harvest in month j

VegGrown: a 1-dimensional array of binary variables; this specifies if vegetable i is chosen to be planted

2.2.3 Objective

The objective function maximizes the overall profits for the farmer. The first part maximizes the amount of revenue generated from crop harvesting; while the second part minimizes the total costs from seed purchase.

Max $\left\{ \left[\sum_{i=1}^{NumVeg} \sum_{j=1}^{NumMonths} RevPerTon_i \times YieldPerMu_i \times z_{ij} \right] - \left[\sum_{i=1}^{NumVeg} \sum_{j=1}^{NumMonths} x_{ij} \times Cost_i \right] \right\}$

2.2.4 Constraints

2.2.4.1 Planting season constraint

This indicates that the area of land for planting during a non-planting season is 0.

$x_{ij} = 0, j \notin PlantingSeason_i, \forall\ i \in \{1, \ldots, NumVeg\}$

2.2.4.2 Planting land constraint

This indicates that the total amount of vegetables planted must be less than the amount of available land throughout the farming months. The original land space is specified by the farmer.

$$Landspace = y_1$$

$$\sum_{i=1}^{NumVeg} x_{ij} \le y_j \ , \ \forall\, j \in \{1, \dots, NumMonths\}$$

The land you have next month will equal to (1) the excess land you have this month, (2) minus the amount of land you used for planting this month, (3) plus the amount of land you gained from harvesting this month

$$y_{j+1} = y_j - \sum_{i=1}^{NumVeg} x_{ij} + \sum_{i=1}^{NumVeg} z_{ij} \ , \forall\, j \in \{1, \dots, NumMonths - 1\}$$

2.2.4.3 Minimum variety constraints

These constraints take into the account the minimum crop variety that needs to be planted and harvested as specified by farmers. Often, farmers who sell vegetables on their own e-commerce platforms need to provide a minimum variety to ensure customer satisfaction. Another reason is to reduce risk by not putting all their eggs in one basket (or crop), in case of factors like diseases or weather conditions that could wipe out certain crops.

$$\sum_{i=1}^{NumVeg} VegGrown_i \ \ge minCropVariety$$

$$VegGrown_i * M \ \ge \sum_{j=1}^{NumMonths} x_{ij} \, , \forall\, i \in \{1, \dots, NumVeg\}$$

(Note M is a very large number that the RHS can never attain)

$$VegGrown_i \ \le \sum_{j=1}^{NumMonths} x_{ij} \, , \forall\, i \in \{1, \dots, NumVeg\}$$

2.2.4.4 Maximum crop quantity constraint

This constraint ensures that one crop type will not dominate by restricting the maximum quantity via maximum proportion of each crop for each month. This also helps to ensure that the model will not just plant a few token vegetables to satisfy the variety constraint, thus ensuring meaningful representation across vegetables planted. The maximum proportion will be adjusted according to the minimum crop variety required to ensure suitable allowance.

$$x_{ij} \le \frac{1}{\sqrt{minCropVariety}} * \sum_{k=1}^{NumVeg} x_{kj} \, , \forall\, i \in \{1, \dots, NumVeg\}, \forall\, j \in \{1, \dots, NumMonths\}$$

2.2.4.5 Harvest constraint

Let $GrowthDuration_i$ be the number of months required for crop i to mature.

This constraint ensures that vegetable cannot be harvested pre-maturely.

$$z_{ij} \ = 0 \ , \forall\, j \in \{1, \dots, GrowthDuration_i - 1\}, \forall\, i \in \{1, \dots, NumVeg\}$$

This constraint ensures that all amount of vegetable is harvest immediately upon maturity.

$z_{i,HarvestTime} = x_{ij}, \forall j \in \{1, \dots, NumMonths\}, \forall i \in \{1, \dots, NumVeg\}$, where HarvestTime = j + $GrowthDuration_i$ - 1

2.2.4.6 Labour constraint

This indicates that the sum of labour required for planting and harvesting is less than the total labour hours available in the farm.

$\sum_{i=1}^{NumVeg} x_{ij} * PlantLabour_i + z_{ij} * HarvestLabour_i \leq LabourAvailable_i \; \forall\, j \in \{1, \dots, NumMonths\}$

2.2.4.7 Budget Distribution constraint

This constraint ensures that the budget distribution prescribed in the stochastic model is adhered to. It limits the budget allocated to each crop category by the budget allocation of the category.

$\sum_{i \in CropCat_k} \sum_{j=1}^{NumMonths} x_{ij} * CropCostPerMu_i \leq BudgetDistribution_k \; \forall\, k \in \{1, 2, 3, 4\}$, *where* $CropCat_k$ *is the set containing all crops belonging to stochastic category k*

3. Results

The Stochastic model returns results for how much budget is allocated to each chosen crop category to minimize risk at 10% CVAR over all crop categories. Total budget used is also shown.

```
Budget allocation to Crop Cat1 = 741645.9571054179
Budget allocation to Crop Cat2 = 0.0
Budget allocation to Crop Cat3 = 0.0
Budget allocation to Crop Cat4 = 18954.04289458212
Amount of budget used: 760600.0
```

Next, the Scheduling model generates results for the optimal planting schedules and harvesting schedules for each vegetable based on farm-specified inputs from the farmer. The model also prints the maximum profit gained for the farm.

In this example, the input entered was to maximize profits for a farm of size 110 mu for a period of 16 months starting from March with 5500 regular labour hours available each month. Minimum crop variety was set to be 7.

```
Solution is Optimal
PLANTING PLAN:
Chinese Cabbage [41.0, 0.0, 0.0, 0.0, 0.0, 12.0, 20.0, 17.0, 12.0, 14.0, 12.0, 13.0, 17.0, 0.0, 0.0, 0.0]
Red amaranth [28.0, 0.0, 38.0, 0.0, 14.0, 9.0, 0.0, 0.0, 0.0, 0.0, 0.0, 10.0, 0.0, 0.0, 13.0, 0.0]
Celery [0.0, 0.0, 0.0, 0.0, 0.0, 12.0, 0.0, 1.0, 20.0, 21.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0]
Leek [0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 20.0, 17.0, 20.0, 21.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0]
Chilli [0.0, 0.0, 39.0, 0.0, 0.0, 0.0, 0.0, 0.0, 1.0, 0.0, 8.0, 0.0, 17.0, 0.0, 0.0, 0.0]
Luffa [0.0, 0.0, 4.0, 0.0, 14.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 1.0, 21.0, 0.0, 0.0]
Hairy Melon [0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 13.0, 0.0]
Gourd Melon [0.0, 0.0, 29.0, 0.0, 10.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 21.0, 9.0, 0.0]
Cherry Radish [41.0, 0.0, 0.0, 0.0, 0.0, 0.0, 13.0, 10.0, 0.0, 0.0, 12.0, 13.0, 10.0, 15.0, 0.0, 0.0]

HARVESTING PLAN:
Chinese Cabbage [0.0, 41.0, 0.0, 0.0, 0.0, 0.0, 12.0, 20.0, 17.0, 12.0, 14.0, 12.0, 13.0, 17.0, 0.0, 0.0]
Red amaranth [0.0, 28.0, 0.0, 38.0, 0.0, 14.0, 9.0, 0.0, 0.0, 0.0, 0.0, 0.0, 10.0, 0.0, 0.0, 13.0]
Celery [0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 12.0, 0.0, 1.0, 20.0, 21.0, 0.0, 0.0, 0.0]
Leek [0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 20.0, 17.0, 20.0, 21.0, 0.0, 0.0, 0.0, 0.0, 0.0]
Chilli [0.0, 0.0, 0.0, 0.0, 0.0, 39.0, 0.0, 0.0, 0.0, 0.0, 0.0, 1.0, 0.0, 8.0, 0.0, 17.0]
Luffa [0.0, 0.0, 0.0, 0.0, 4.0, 0.0, 14.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 1.0, 21.0]
Hairy Melon [0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0]
Gourd Melon [0.0, 0.0, 0.0, 0.0, 29.0, 0.0, 10.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 21.0]
Cherry Radish [0.0, 41.0, 0.0, 0.0, 0.0, 0.0, 0.0, 13.0, 10.0, 0.0, 0.0, 12.0, 13.0, 10.0, 15.0, 0.0]

Labour Hours used:
[1031.0, 2034.0, 1662.0, 646.0, 1056.0, 1120.0, 1339.0, 1459.0, 1576.0, 1151.0, 1058.0, 1148.0, 1662.0, 1357.0, 758.0, 1232.0]

Cost:
[20826.2, 0.0, 22630.1, 0.0, 7424.3, 5290.44, 9816.55, 8255.61, 8527.25, 9008.79, 6352.28, 6777.79, 8713.14, 12939.4, 8229.77, 0.0]

Maximised Profit: 2.5092459135922015e7
```

<u>Interpretation of results:</u> As shown, the model suggests the farmer to plant 4 Mu of Luffa in May (Month 3), which can be harvested at the end of its growth duration in July (Month 5). The same crop is again planted again in July, March and April of the subsequent year. The model also suggests the labour required per month, allowing the farmer to quickly identify the labour capacity. All in all, the maximum profit gained in this 16-months farming cycle is ¥25,092,459.

4. Discussion

4.1 Limitations

We have built a basic model using Linear Programming techniques to derive an optimal land allocation for 30 vegetable crops, based on several crucial constraints and farm conditions. However, some of the assumptions upon which the model is based on may not hold true in practice:

<u>Lack in Consideration for Partial Harvesting</u>

The model dictates that all planted vegetables must be harvested all at once when their growth durations are up. However, in reality some vegetables can bear fruits more than one time such that there are multiple harvest periods throughout its lifespan even after a single planting period. Hence, by harvesting only once and using the land afterwards to plant another vegetable may not be cost-effective for farmers. To mitigate this, research needs to be done on harvest schedules of different vegetables and include it separately as another input in the model and allow for additional constraints which determine when the vegetable can be harvested again.

Lack in Consideration for Expressing Different Land Units

Based on the current model, vegetables can only be planted and harvested in units of 1 mu, which is approximately 0.165 acres or 667 m^2. However, depending on the size of the farm and whether it is highly commercialised or a simpler family business serving the local community, farmers may choose to plant in smaller planting units. To cater to greater variety in size of farms, the model can be modified to allow users to choose their preferred unit of measurement for land size. This allows farmers to be more precise about exactly how much they wish to plant, and not comply them to plant at least 1 mu of each chosen vegetable by default.

4.2 Suggestions for Improvements

The existing model can be further refined by considering conditions in reality and integrating other cost-profit structures unique to different farms. Besides the constraints and variables included by Section 4.1, the following recommendations can boost robustness of the model:

Input more demand and supply factors over time

A key assumption made is that all harvested plants can be sold in the market and hence revenue is guaranteed. In reality, this is not the case as demand changes according to season and consumer preferences. Additionally, vegetable yield only captures productivity of two factors of production: land and labour. Productivity of capital and entrepreneurship is not quantified. While entrepreneurship is difficult to model, capital productivity is easier to measure. The use of machines is becoming more popular in labour-intensive spaces like agriculture. Mr Hu may invest in state-of-the-art machines in the future to improve farming processes and raise land and labour productivity. Therefore, the model should always update its demand and supply constraints to ensure true reflection of farming conditions.

Specify locations of crops

The model can be made more meaningful if it specifies the planting location of crops on the farm. Common farming practices such as crop rotation and companion planting can create growth synergies or resource savings and thereby affecting the crop yield. For instance, some plants are natural pollinator attractors, and their presence can attract pollinators such as bees and other insects which can assist in pollinating majority of the farm. One possible way to achieve this effect in our model is through a 3D array [with 2-dimensions representing the plots of land and 1-dimension representing the plants planted in that plot]. Constraints are then added to consider the surrounding plots before determining the yield of the plant, similar to the discrete optimization problems covered in the lecture. This would help the farmer optimize both locations and time to plant the different plants.

5. Conclusion

Technology will be the key enabler to allow China to produce enough food for the growing population in a safe and healthy way. Our Lili model serves as pivotal starting point for enabling more optimised crop mix decisions with regards to planting and harvesting, and can potentially be expanded to factor in other key activities on the farm. We hope to increase the farm revenues for Mr. Hu and the millions of small-sized family farms and to improve China’s rural agricultural landscape!

让中国 1000 万家庭吃上实惠的放心菜! ☺

Appendix A – Background of 简法农庄

Our Lili 粒粒 model is inspired by a farm in Hangzhou, China. Here are some pictures taken at the farm for your viewing pleasure:











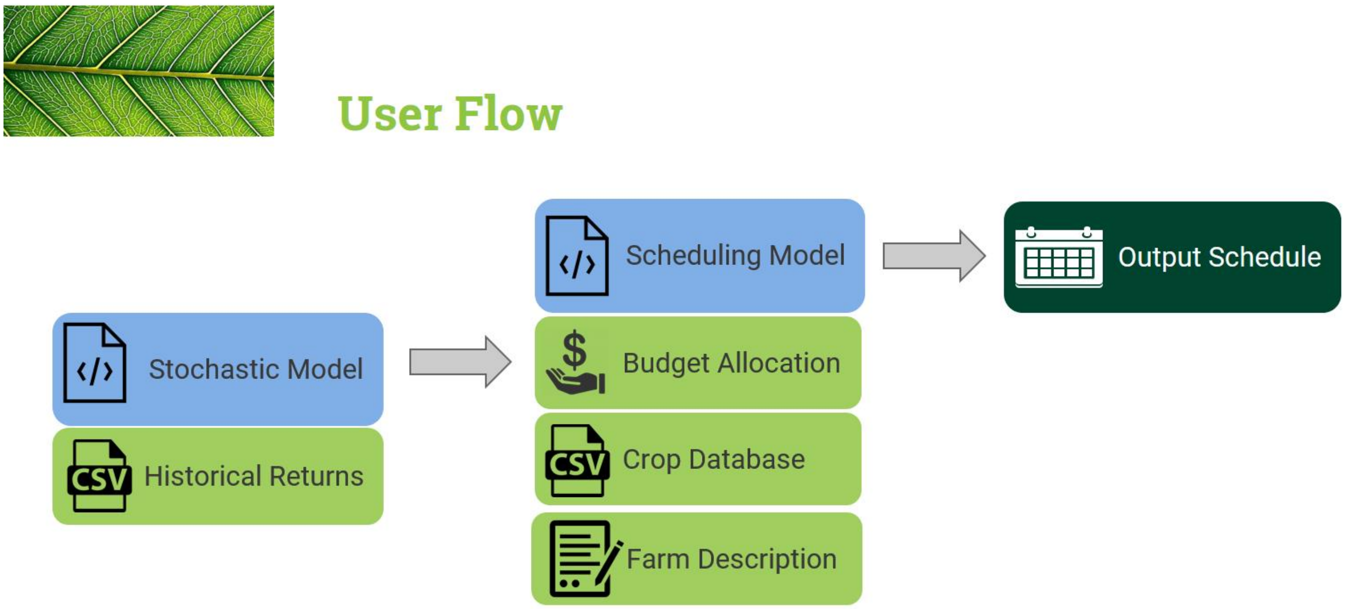
Appendix B – List of Vegetable Crops

The 30 vegetable crops along with their categories are shown below:

Category 1 白菜 Chinese Cabbage 红苋菜 Red amaranth 西芹 Celery 韭菜 Leek 洋葱 Onion 节瓜(毛瓜)Hairy Melon 葫芦瓜 Gourd Melon 胡萝卜 Carrot 樱桃萝卜 Cherry Radish	**Category 2** 小叶茼蒿 Crysanthemum 菠菜 Spinach 甜椒 Capsicum 番茄 Tomato 白萝卜 Radish 姜 Ginger
Category 3 白花菜 Cauliflower 薯仔/马铃薯 Potato	**Category 4** 芥兰 Kale 西兰花 Broccoli 菜心 Choy Sum 葱 Spring Onion 蒜頭 Garlic 辣椒 Chilli 丝瓜 Luffa 冬瓜 Winter Melon 苦瓜 Bitter Gourd 意大利青瓜 Zucchini 黄瓜 Cucumber 南瓜 Pumpkin 番薯 Sweet Potato

Appendix C – User Flow of Model

A diagrammatic representation of how the Stochastic and Scheduling model work in conjunction is shown below:



Appendix D – Definition of Variables

Stochastic Model Variables

DEFINITIONS

a, alpha (%)
Conditional Value at Risk(average of the losses that occur beyond the VaR cutoff point in the distribution)

expReward
Amount of RMB expected by the Farmer

RetMean
Means of the returns (in the historical data)

Budget *(RMB)*
Amount of RMB available for use by the Farmer

T
Periods of historical returns available

RetTable
Table of the returns (the historical data)

Scheduling Model Database Variable

CROP DATABASE

Yield *(Ton/Mu)*
Ton of vegetable harvested from 1 Mu (亩)

Revenue *(RMB/Ton)*
Amount of RMB received per ton of vegetable sold

Growth Duration *(mth)*
Number of months to maturity

Planting Season *(mth)*
Months that vegetable can be planted

Labour *(Hr/Mu)*
Man-hours required to plant and harvest 1 Mu of vegetable

Cost *(RMB/Mu)*
Cost required to plant and harvest 1 Mu of vegetable

Category
Stochastic category that vegetable belongs to

Scheduling Model Input Variables

FARM DESCRIPTION

Start Month *(mth)*
Month (from Jan to Dec) where farmer begins farming

Farming Duration *(mth)*
Total months of farming

Labour *(Hrs/mth)*
Man-hours available per month

Farm Size *(Mu)*
Total size of arable land

Minimum Variety
Number of vegetable variety required.