SUSTAINABLE BIOECONOMY
FOR ARID REGIONS

# Biomass Supply Chain and Transportation:

# Instruction Manual


Systems and Industrial Engineering
Department

*By*

Daniel A. Zuniga Vazquez, Shunyu Yao, Neng Fan


July 23, 2022

Acknowledgments

This material is based upon funding provided by the

USDA-NIFA, Grant # 2017-68005-26867.

Any opinions, findings, conclusions, or recommendations expressed in this publication/work are those of the authors and do not necessarily reflect the view of the U.S. Department of Agriculture. All authors state that there is no conflict of interest.

The authors would also like to acknowledge valuable input and assistance from the following individuals and organizations who kindly provided expertise, data, and/or feedback during the development of this instruction manual:

- Amy E. Landis (Colorado School of Mines)
- Blase Evancho (University of Arizona)
- Clark Seavert (Oregon State University)
- Evan Sproul (Colorado State University)
- Hailey M. Summers (Colorado State University)
- Jason C. Quinn (Colorado State University)
- Paul Gutierrez (New Mexico State University)
- Sita Khanal (New Mexico State University)
- Trent Teegerstrom (University of Arizona)
- VeeAnder Mealing (Colorado School of Mines)

All the data and source code in this manual can be found in the Github website: https://github.com/omegayao/SBAR.

# Contents

# List of Figures

# List of Tables

Chapter 1

# Introduction

## 1.1 Overview

The Biomass Supply Chain and Transportation (BSCT) optimization tool has been developed by the Sustainable Bioeconomy for Arid Regions (SBAR) project in support with the U.S. Department of Agriculture (USDA) to evaluate scenarios for crop adoption and their scaled-up to profitable production.

For the economical, environmental, and social impacts analyses of the biomass supply chains for the adopted crops, BSCT uses stochastic scenarios to capture the uncertainties of the adoption rates of each field. Afterward, a stochastic optimization is deployed to identify optimal decisions for facility locations, transportations from fields to facilities, and finally to customers, with a multi-objective approach to quantify the economic benefits (minimizing the costs of supply chains), environmental impacts (minimizing $CO_2$ equivalent greenhouse gas emissions), and social impacts (maximizing the local accrued jobs). A Geographic Information System is used for capturing field information and relevant factors, as well as identifying potential processing facility locations.

Once the optimal processing facility location and supplying farms have been identified, BSCT can perform a detailed analysis of the farm production planning and machinery scheduling can through its optimization model for perennial and non-perennial crops that maximizes the farmers' net present value. Such optimization model allows for multi-crop production planning, multi-machinery scheduling, and crop rotation. To address the water scarcity problem in arid and semi-arid regions, perennial and nonperennial low-water-use crops can be considered in the crop rotation. Based on the Geographic Information System, the optimal scheduling routes are determined and the irrigation water requirements are analyzed.

Fig. 1.1 presents the overall structure of BSCT. Step 1 focuses on defining the sets and input parameters for the integration of the economical, environmental, and social impacts into the design of the biomass supply chains. Step 2 identified the optimal processing facility locations and the expected production from the supplying farms based on the optimization model. Step 3 defines the sets and input parameters for the production planning and machinery scheduling. Step 4 identifies the optimal production planning with crop rotation and multi-machinery scheduling through the optimization model. Two test scenarios are provided for the Steps 1 and 2 and two test scenarios are provided for Steps 3 and 4 as follows:

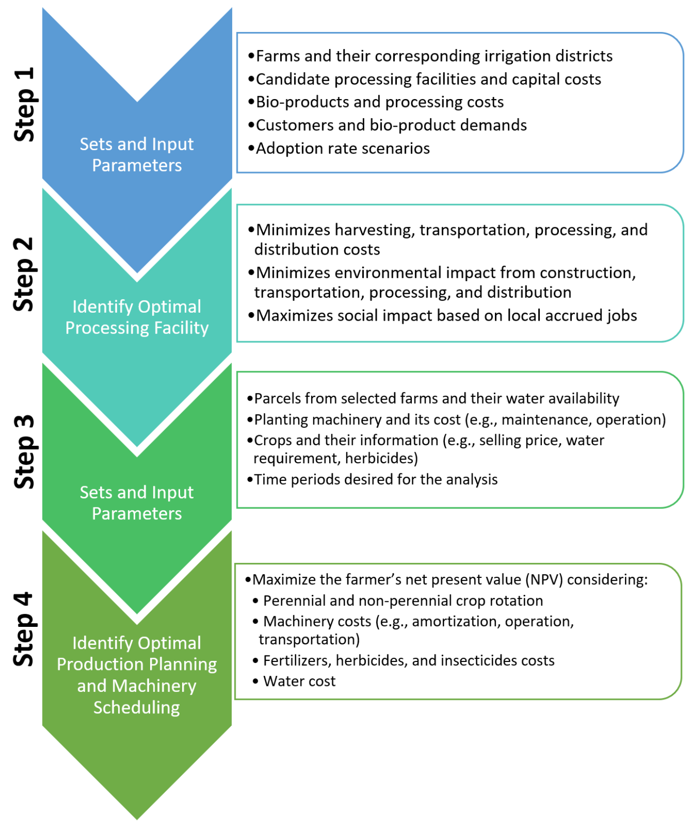


Figure 1.1: BSCT overall structure

- Guayule optimal processing facility location and expected farm production for Pinal and Maricopa counties in Arizona (refer to [1]).
- Guar optimal processing facility location and expected farm production for Dona Ana County in New Mexico (refer to [1]).
- Guayule optimal production planning and multi-machinery scheduling for Pinal County in Arizona (refer to [2]).
- Guar optimal production planning and multi-machinery scheduling for Cochran and Hockley counties in Texas (refer to [2]).

- Guayule optimal harvest planning and machinery scheduling for Pinal County in Arizona (refer to [3]).

## 1.2 Need for a sustainable bio-economy for arid regions

Increasing competition in the agricultural sector is causing water to be a more scarce resource. In the U.S., approximately 80% of the consumptive water use is accounted for agriculture and it reaches up to 90% in the Western States [4]. Furthermore, frequent droughts and climatic change also increase the pressure for limiting the agricultural sector's water use [5]. The situation worsens in arid and semi-arid regions with low rainfall and high evaporation [6]. Therefore, ensuring efficient water use is an urgent problem to be solved [7]. To address this issue, farmers can gradually adopt low-water-use crops such as guar and guayule. These two crops are starting to create an interest for researchers and industries as a biomass resource [8].

Guayule and guar are low-water-use, drought-tolerant, and heat-resistant crops. Guayule is native to the Chihuahuan Desert, from southwestern U.S. regions to the northern Mexico, and it has been often treated as a source of natural rubber. Due to emerging economies in countries around the world, the demand for natural rubber has proliferated, and the guayule's natural rubber has been found to be an attractive commercial option as a domestic rubber source [9]. Guayule has even been declared as "a critical agricultural material in times of crisis" [10]. Similarly, guar is an important legume native from India and the source of guar gum. Partially due to its stabilizing and thickening properties, guar gum and its derivatives are essential in numerous applications such as paints, batteries, pharmaceuticals, ceramics, food, petroleum, cosmetics, textiles, and agriculture [11, 12]. BSCT is developed as an effort to achieve a scale-up to profitable production for guayule and guar.

## 1.3 Purpose of Report

BSCT will enable farmers and researchers assess the adoption of crops based economical, environmental, and social impact analysis. BSCT will also enable farmers evaluate the economical feasibility production planning and scheduling programs of the adopted crops. For such purpose, this instruction manual shows a step by step on the know-how to use BSCT and understand its results.

## 1.4 Required Software and Installation Guides

The following software is required by BSCT. Note that a license is required in several cases, while other are available for free. Download and detailed information links are provided for each case.

- **ArcGIS Pro**
  ArcGIS Pro is a desktop Geographical Information System (GIS) software

developed by Esri.
Related links:
https://www.esri.com/en-us/arcgis/products/arcgis-pro/resources

- **Duo Mobile**
  Secure access to work and personal, cloud and on-premises apps with one simple app.
  Related links:
  https://duo.com/product/multi-factor-authentication-mfa/duo-mobile-app
  https://it.arizona.edu/documentation/install-duo-mobile-app-netid-plus#duo

- **Cisco AnyConnect VPN**
  Secure virtual private network (VPN) access.
  Related links:
  https://www.cisco.com/c/en/us/products/security/anyconnect-secure-mobility-client/index.html
  https://it.arizona.edu/service/ua-virtual-private-network-vpn

- **PuTTY**
  PuTTY is a free security shell (SSH) protocol and telnet client for Windows.
  Related link:
  https://www.putty.org/

- **WinSCP**
  WinSCP is a free SSH File Transfer Protocol (SFTP) and File Transfer Protocol (FTP) client for Windows.
  Related link:
  https://winscp.net/eng/index.php

- **High Performance Computing (HPC)**
  For an efficient solution of the large-scale models from BSCT, the use of a HPC is recommended. A service overview and guide for the University of Arizona HPC is provided in the following links:
  Service overview: https://it.arizona.edu/service/high-performance-computing
  User Guide: https://public.confluence.arizona.edu/display/UAHPC/User+Guide

- **CPLEX** IBM ILOG CPLEX Optimization Studio is an optimization software package.
  Related link:
  https://www.ibm.com/analytics/cplex-optimizer

- **Gurobi**
  Gurobi Optimizer is a state-of-the-art solver for mathematical programming.
  Related link:
  https://www.gurobi.com/

## Chapter 2

# Integrating Environmental and Social Impacts into Optimal Design of Guayule and Guar Supply Chains

This research proposes a stochastic optimization to make optimal decisions for facility locations, transportations from fields to facilities, and finally to customers, with a multi-objective approach to quantify the economic benefits, environmental impacts, and social impacts. The adoption rates are integrated into the model as a random variable which allows the capture of the farmers' uncertainty to adopt guayule and guar in their fields. The quantified economic benefits include harvesting costs, transportation and distribution costs, production costs, and construction costs for the processing facilities. Environmental impacts are accounted for through $CO_2$ equivalent GHG emissions cost of the harvesting, production, transportation, distribution, and construction processes. Social impacts are quantified based on the local accrued jobs per county for harvesting, transportation, production, distribution, and construction. Additionally, a GIS analysis is integrated into the model's input parameters to identify potential candidates for processing facility location. Irrigation water requirements are analyzed through GIS irrigation district layers and their water availability. Independent weighted factors are considered within the quantification of the economic benefits, environmental impacts, and social impacts for sensitivity analysis purposes. This allows an informed decision making based on assigning higher weights on one or multiple economic, environmental, and social impacts as required. The proposed approaches are evaluated based on the cases of two areas: the Maricopa and Pinal counties in Arizona for the guayule supply chain, and the Dona Ana County in New Mexico for the guar supply chain. For full access to the model and input parameters, refer to [1].

## 2.1 Model Structure

To model the guayule and guar supply chains for the American Southwest, several economic, environmental, and social inputs are required as well as geographical information as shown in Fig. 2.1. To assess the economic impact, the inputs considered in the analysis include capital construction, harvesting, hauling, loading, unloading, transportation, processing, and distribution costs. Similarly, the environmental impact is evaluated through the equivalent greenhouse gas emissions of construction, transportation, processing, and distribution. As per the social impact, it is determined based on the local accrued jobs and the earnings per work hour.

Based on the number of farms, their acreage, the sales as defined by the Census of Agriculture 2017 [13], and discussions with industry experts, the farm groups cultivating cotton, grains, and oilseeds (CGO) crops are identified as the potential candidates to adopt guar and guayule [8]. The information of the farm groups such as the specific type of crop, the geographical location, and the acreage is obtained from USDA CropScape [14]. Candidates for processing facilities locations are identified using GIS layers for transportation, development, water resources, and the farm group's information. Through discussion with SBAR researchers, a 15% of adoption rate from the overall crops is defined as a baseline for the stochastic scenarios. The optimal routes are determined using the origin-destination cost matrix function of ArcGIS Pro 2.4.0.

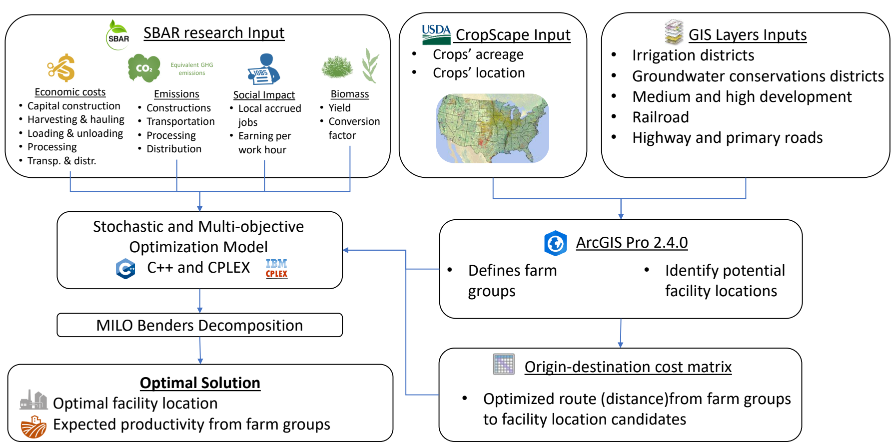


Figure 2.1: Biomass supply chain optimal design: model methodology

## 2.2 GIS Data and Geoprocessing Tools

### 2.2.1 Importing Data

The crops and farms information is obtained from USDA CropScape [14] as shown in Fig. 2.2. To download the information, select the polygon tool indicated in Step 1 and select the area of interest as marked in Step 2. In this example, New Mexico State is selected as the area of interest. Afterward, click on the download button indicated in Step 3 and a pop-up window will appear where you can select the years for which the information is desired (Step 4). The information will be downloaded in a raster format file.

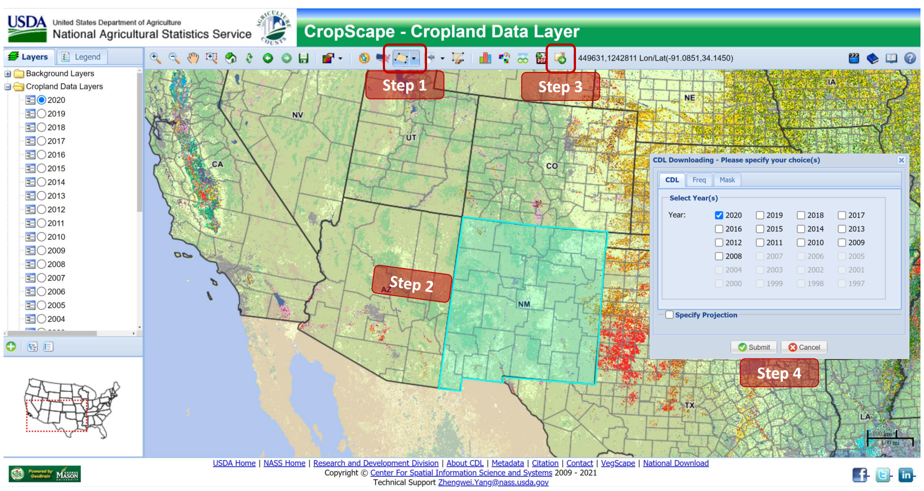


Figure 2.2: CropScape information

The next step is to import the raster file into ArcGIS Pro. For such purpose, please refer to Fig. 2.3. As shown in Fig. 2.3 a), to add the raster file, go to the Map menu, click on the Add Data drop down menu, and select the Data tool. You will need to provide the location of the saved raster file from CropScape. Once the raster file is uploaded, it will be overlaid on the map as presented in Fig. 2.3 b). From the Appearance menu (1), select the Symbology tool and the menu will appear on the right-hand side (2). The Symbology tool provides the crops' name and legend information. However, the raster file needs to be transformed to a fully editable version before proceeding with the analysis, in this case, the selected format is a polygon layer. The transformation process is presented in Fig. 2.3 c). First, select the Analysis menu (1), search for the Raster to Polygon tool (2), and complete the required information. Make sure to check the "Create Multipart Features" box to maintain the crop classification during the transformation. Finally select Run (3) to obtain the polygon layer. Fig. 2.3 d) shows an example where once the polygon layer was created, all crops where removed except CGO. Note how the Symbology changes, and the classification is defined by gridcode (1) and the detailed information each polygon is also provided based on the gridcode (2).

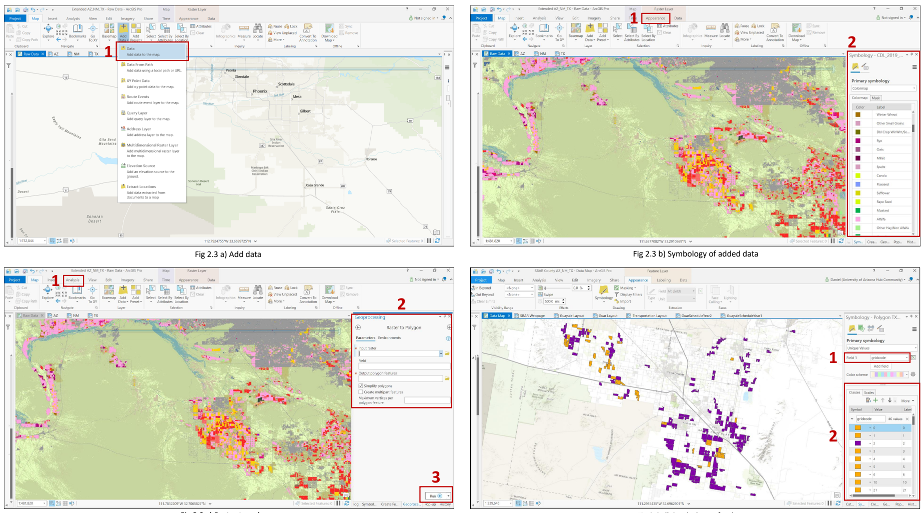

Fig 2.3 a) Add data

Fig 2.3 b) Symbology of added data

Fig 2.3 c) Raster to polygon

Fig 2.3 d) Symbology of polygons

Figure 2.3: GIS data import

### 2.2.2 Customization of GIS Data

Once the polygon layer has been created and only the polygons from the farms of interest remain, the next step is to customize the layer as shown in Fig. 2.4. As shown in Fig. 2.4 a), more than one color range can be created for the polygon layer. For such purpose, right-click on the Map, select the Attribute Table, and it will appear on the bottom part of the screen (1). To add new attributes, select the button Add (2) and specify the characteristics for each attribute. In this example, three different color attributes are created (3) to ease switching between the color ranges for each polygon (farm). The specific color range is defined in the Symbology pane (4). Similarly, additional attributes need to be added such as the acreage and the geographic location of the polygon (farm) location as shown in Fig. 2.4 b). In this example, the attributes AreaAcre, X_Longitude, and Y_Latitude were created (1).

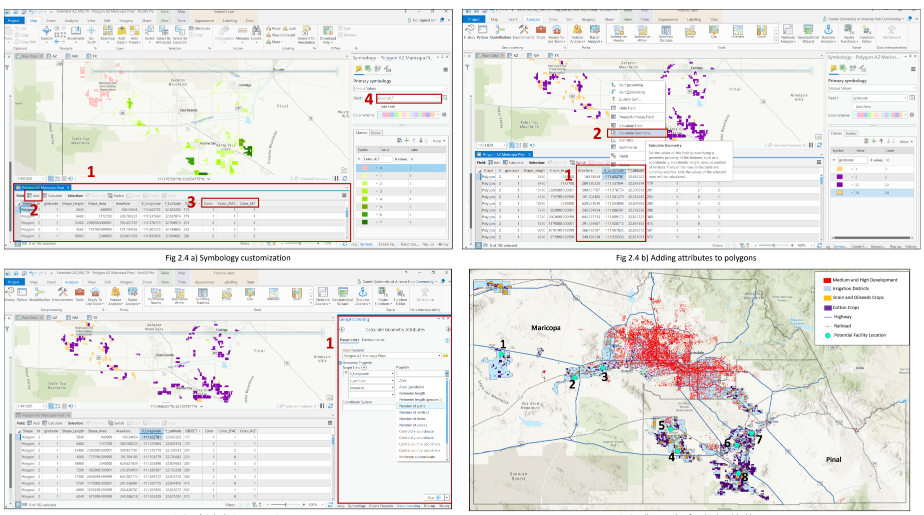


Figure 2.4: GIS data customization

The attributes will be created empty and the tool Calculate Geometry (2) needs to be selected to calculate each one. Fig. 2.4 c) shows the geoprocessing tool of Calculate Geometry Attributes (1), where you just need to assign the attributes of interest in the target fields and specify what property will be calculated for each attribute. Fig. 2.4 d) provides an example where additional layers have been added including main roads, development areas, and irrigation districts. Additional layers can be directly imported following the steps in Fig. 2.3 a).

### 2.2.3 Distance Matrix and Printing Formats

After customizing the polygon layer, the next information to calculate is the distance matrix (Fig. 2.5), i.e., the distance between origins and locations, or in this case, farms and candidate processing facilities. Fig. 2.5 a) shows how to access the Origin-Destination Cost Matrix through the Analysis menu. Once selected, a new and empty Cost Matrix layer will be created. To add the origins and destinations, the tools indicated in Fig. 2.5 b), i.e., Import Origins and Import Destinations (1) can be used. After imported, they will appear in the content pane under the created Cost Matrix layer (2). A transportation layer can be selected to determine the distances or use the ESRI extensive transportation information (e.g., streets, major and minor roads, and highways) by default. After importing the origins and destinations, click the Run button (1).

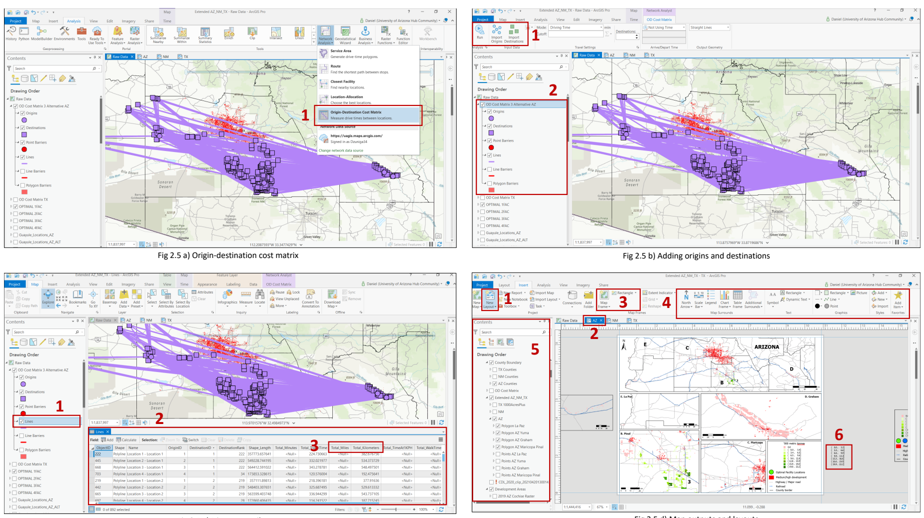


Figure 2.5: GIS distance matrix and map output

Results will appear in form of straight lines as presented in Fig. 2.5 c) from each origin to every destination. To see detailed information on the results, right-click the line checkbox (1) and open its Attribute table (2). The Attribute table will provide the distance in kilometers or miles. Note that the straight lines only indicate relationship and the calculated distance is bases on the transportation layer, i.e., the roads and highways. Finally, to create print the results, a map layout can be created as shown in Fig. 2.5 d). To do so, select the Inset Menu and click on the New Layout (1). Once a layout has been selected, you can navigate back to the previous map using the tabs indicated in (2). Multiple frames can be added to a single map layout to show multiple maps together (3). The example provided includes 5 different map frames. Additional information such as legends, scale bars, and north arrows can be automatically created (4) if desired. If a layer is not being displayed, make sure its checkbox is marked in the Content Pane (5). Backup information or alternative map frames can be stored outside of the map layer (6) for ease access in future analysis.

## 2.3 Defined Parameters

The sets and indices are presented in Table 2.1 and Table 2.2 defines the required parameters for planning, harvesting, production, transportation, and distribution, as well as for the environmental and social impacts. The abbreviations of the units are defined as follows: acre (ac), $CO_2$ equivalent ($CO_2$ eq), cubic meter ($m^3$), hour (hr), kilometer (km), U.S. dollar (\$), million U.S. dollar (M\$), metric tonne (t), and thousand metric tonne (Kt).

Table 2.1: Biomass Supply Chain Optimal Design: Sets and Indices

| | |
|---|---|
| $\mathcal{I}$ | Set of irrigation districts, indexed by $i$. |
| $\mathcal{F}$ | Set of farms, indexed by $f$. |
| $\mathcal{F}^i$ | Set of farms in water irrigation district $i$. |
| $\mathcal{B}$ | Set of processing facilities, indexed by $b$. |
| $\mathcal{P}$ | Set of bio-products, indexed by $p$. |
| $\mathcal{C}$ | Set of customers, indexed by $c$. |
| $\Omega$ | Set of adoption rate scenarios, indexed by $\omega$. |

Table 2.2: Biomass Supply Chain Optimal Design: Input Parameters

| | |
|---|---|
| $\theta^1, \theta^2, \theta^3$ | Weighted factor for the economic, environmental, and social objectives, respectively. |
| $Prob(\omega)$ | Probability of adoption rate scenario $\omega$. |
| $C_b^{cap}$ | Capital construction costs for processing facility $b$ (\$). |
| $K$ | Number of planning processing facilities. |
| $Max_b$ | Minimum annual capacity of processing facility $b$ (t). |
| $Min_b$ | Maximum annual capacity of processing facility $b$ (t). |
| $C^H$ | Crop's harvesting and hauling cost (\$/t). |
| $C^L$ | Crops's loading and unloading cost (\$/t). |
| $Y_f$ | Yield for crop per acre of farm $f$ (t/ac). |
| $A_f^{(\omega)}$ | Area of supply of farm $f$ for adoption rate $\omega$ (ac). |
| $P_f^{(\omega)}$ | Crop productivity of farm $f$ for adoption rate $\omega$, e.g., $P_f^{(\omega)} = Y_f \times A_f^{(\omega)}$ (t). |
| $W_i$ | Water availability in irrigation district $i$ ($m^3$). |
| $U_f$ | Irrigation water use by farm $f$ ($m^3$/t). |
| $C^P$ | Processing cost of biomass (\$/t). |
| $\theta_p$ | Conversion factor for bio-product $p$. |
| $D_{pc}$ | Demand of bio-product $p$ for customer $c$ (t). |
| $C^T$ | Transportation cost for biomass (\$/t×km). |
| $C_p^D$ | Distribution cost for bio-product $p$ (\$/t). |
| $L_{fb}$ | Distance from farm $f$ to processing facility $b$ (km). |
| $\mu_b$ | Construction emission factor for processing facility $b$ ($CO_2$ eq). |
| $\beta$ | Transportation emission factor ($CO_2$ eq/t×km). |
| $\gamma_p$ | Distribution emission factor for bio-product $p$ ($CO_2$ eq/t×km). |
| $\delta_p$ | Processing emission factor for bio-product $p$ ($CO_2$ eq/t). |
| $\varepsilon$ | Emission price (\$/$CO_2$ eq). |
| $UB^{EnvImp}$ | Emissions impact upper-bound (\$). |
| $E_b^B, E_f^H, E_f^T, E_b^P, E_b^D$ | Earning per work hour for construction of processing facility $b$ harvesting farm $f$, transportation from farm $f$, production of bio-product from processing facility $b$, and distribution of bio-product from processing facility $b$, respectively (\$/hr). |
| $H^T, H_b^D$ | Number of work hours required for transportation and distribution of bio-product from processing facility $b$, respectively (hr/t×km). |
| $H^H, H_b^P$ | Number of work hours required for harvesting and production of bio-product from processing facility $b$, respectively (hr/t). |
| $H^B$ | Number of work hours required for processing facility construction (hr). |
| $LB^{SocImp}$ | Social impact lower-bound (\$). |

### 2.3.1 Identifying processing facility candidates

To identify the processing facility location candidates, a GIS analysis can performed with information from irrigation districts, CGO crops, medium/high development, highway, and railroad layers which are overlapped to identify potential facility locations and farms. The GIS layers such as the development levels, county boundaries, highways, railroads, and irrigation districts are collected from online databases. Distances between farms, processing facility location candidates, and customers are calculated using ArcGIS Pro as shown in Fig. 2.5. The criteria used to identify the processing facility location candidates are as follows:

(a) Accessibility to biomass supplies, close to feasible planting fields.

(b) Within a 2-mile straight line distance to major roads.

(c) Within a 2-mile straight line distance to the railway.

(d) Within a 2-mile straight line distance of the pipeline/major water bodies.

(e) Avoid the developed areas (high intensity and medium intensity areas).

### 2.3.2 CPLEX Parameter File Creation

A Microsoft Excel Macro-Enabled file is provided to created the input parameter file required for the CPLEX optimization software (refer to the file MultiObjective_Data_Original.xlsm) Fig. 2.6 to Fig. 2.8 will guide you through the four steps in the xlsm file. Fig. 2.6 highlights the instruction for the 4 steps, where step 1 is just to provide the desired name for the output TXT file, which will be used as an input for CPLEX. Step 1 additionally requires to fill the cardinality of the sets defined in Table 2.1, e.g., the number of farms, number of irrigation districts, and number of candidate processing facilities. Once all the information in Step 1 has been completed, Step 2 just requires to click the Restructure Table button and the spreadsheet will restructure based on the provided cardinality of the sets as shown in Fig. 2.7.

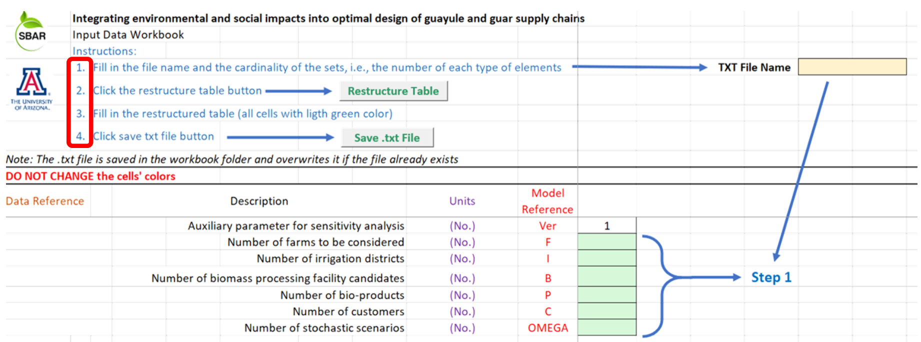


Figure 2.6: Input parameter walkthrough: Step 1

Once all of the information of the restructured table has been completed, i.e., every cell in green color as shown in Fig. 2.8, Step 4 is to click the Save .txt File button and a text file with the assigned name will be created within the same folder of the .xlsm file.

Note: **Do not use the original .xlsm file**. Since the file uses macros for the restructuring, changes cannot be easily undone. It is better to create a copy of the .xlsm file to work in and leave the original as a backup. Additionally, **the color of the cells must not be modified** since the macros are color coded. If an update in a parameter is required (except for the sets cardinalities) he values can be changed directly and the filed saved again with the Save .txt File button to update the text file.

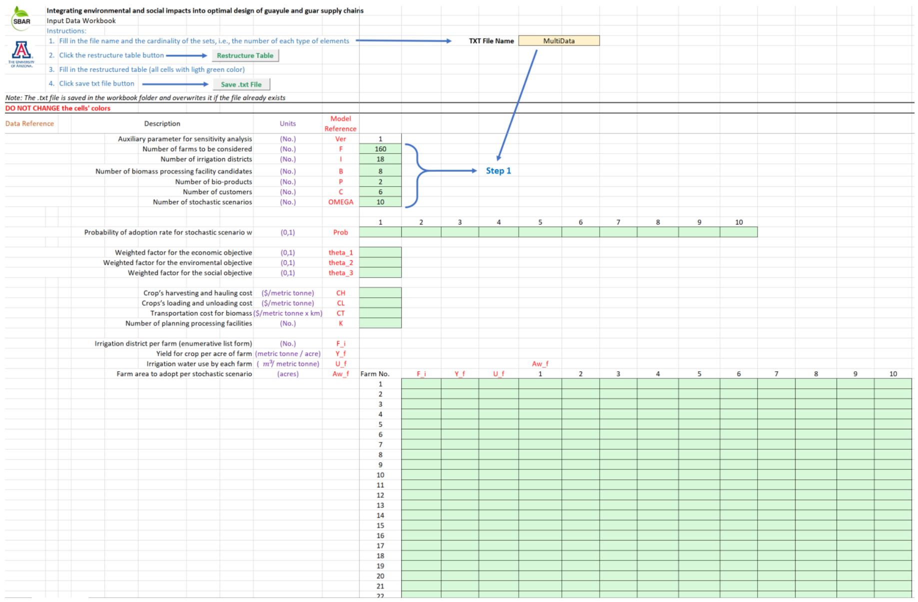


Figure 2.7: Input parameter walkthrough: Step 2

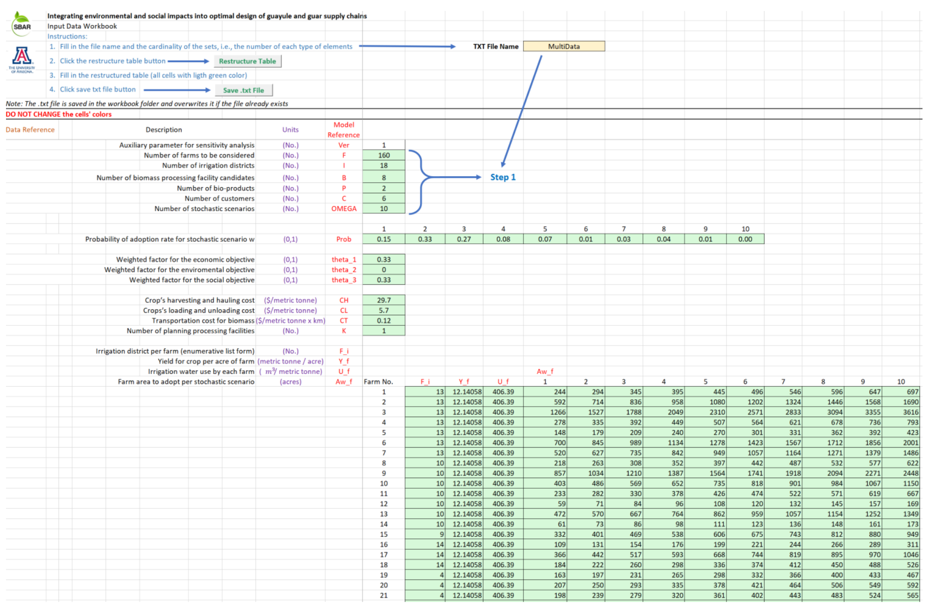


Figure 2.8: Input parameter walkthrough: Steps 3 and 4

## 2.4 Optimization Tools and Methods

### 2.4.1 CPLEX Optimizer Through a High Performance Computer

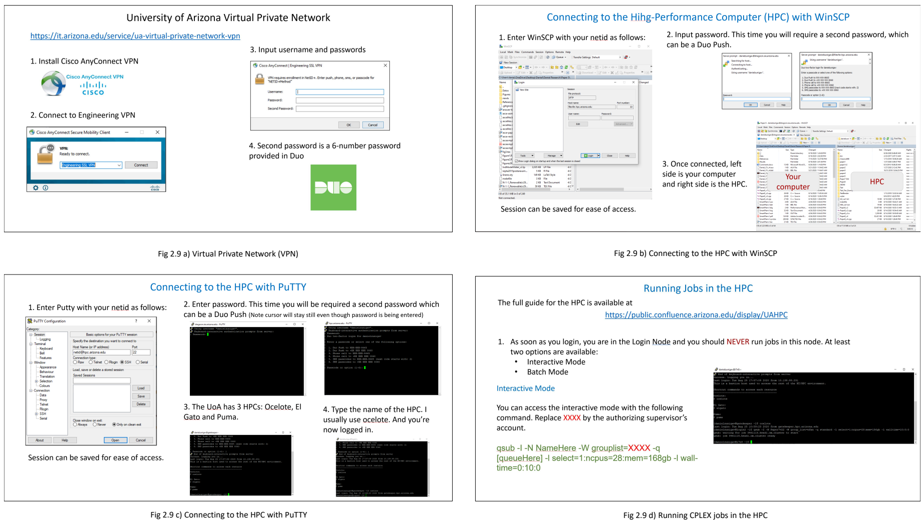


Figure 2.9: Running CPLEX optimizer in the HPC (Part 1)

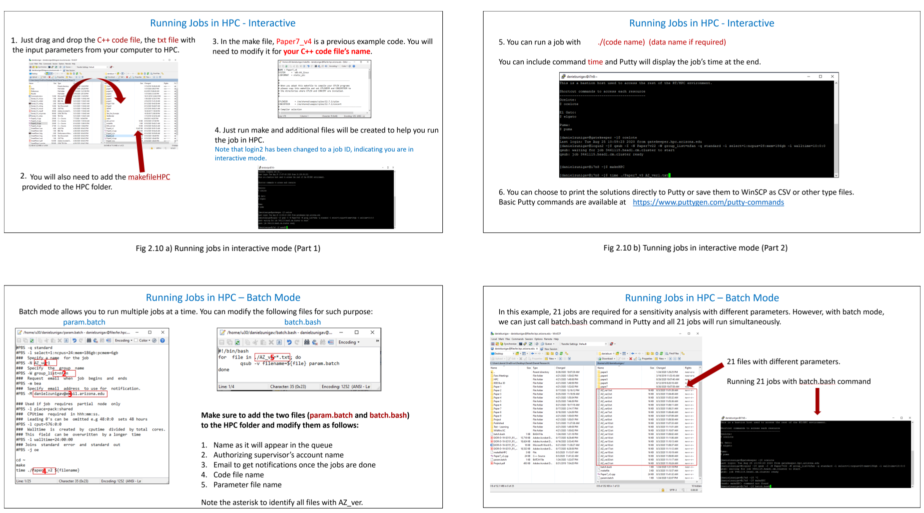


Figure 2.10: Running CPLEX optimizer in the HPC (Part 2)

### 2.4.2 Reporting Outputs

Table 2.3: Biomass Supply Chain Optimal Design: Decision Variables

| | |
|---|---|
| $x_b$ | Binary variable that is 1 if processing facility $b$ is added and 0 otherwise ($\boldsymbol{x}$ is the vector form). |
| $cap_b$ | Capacity of processing facility $b$ ($\boldsymbol{cap}$ is the vector form). |
| $u_{fb}^{(\omega)}$ | Commodity flow of biomass from farm $f$ to processing facility $b$ ($\boldsymbol{u}$ is the vector form) (metric tonnes). |
| $v_{pbc}^{(\omega)}$ | Commodity flow of bio-product $p$ from processing facility $b$ to customer $c$ ($\boldsymbol{v}$ is the vector form) (metric tonnes). |

## 2.5 Model Assumptions

Model assumptions includes that certain information about the probability distribution for the stochastic scenarios for the adoption rate is known beforehand. Additionally, a single crop is considered per farm or farm group and capital costs do not consider interest rates.

## 2.6 Extended Results for the American Southwest

This research an extension of Sections 2.1-2.5 to the American Southwest. To identify potential farm groups for the adoption of guayule and guar, all of the farms groups within this region are evaluated based on their type of crop, location, water availability, and acreage. A 15% base of the overall crops is defined for the stochastic scenarios with log-normal distribution probabilities. This research explores further the contributions from the model of Chapter 2 by considering multiple processing facility location candidates. A sensitivity analysis is performed for multiple optimal facility locations based on the economic, environmental, and social impacts of the guayule and guar supply chains. Three case studies are presented as follows: i) Arizona for guayule adoption (Fig 2.11), ii) New Mexico for the adoption of guar (Fig 2.12), and iii) Texas for guar adoption (Fig 2.13).

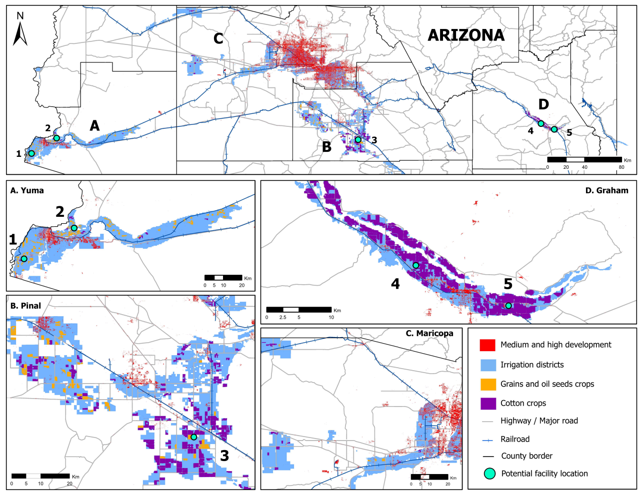


Figure 2.11: Arizona's GIS information

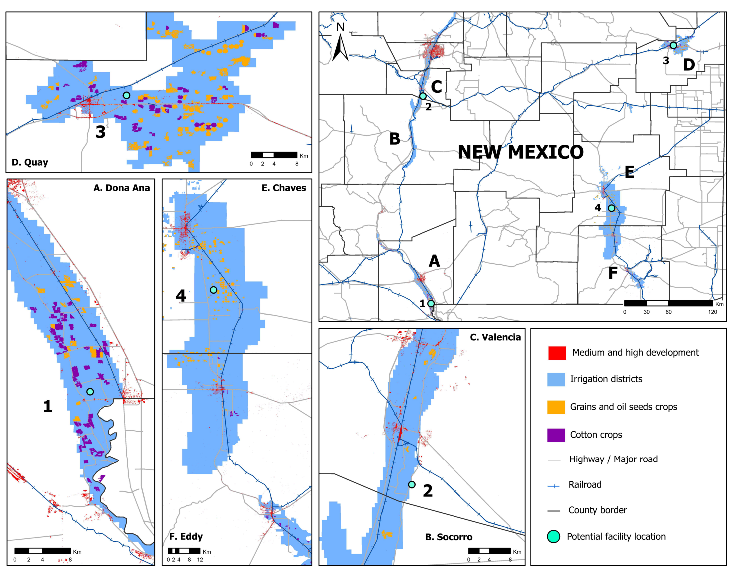


Figure 2.12: New Mexico's GIS information

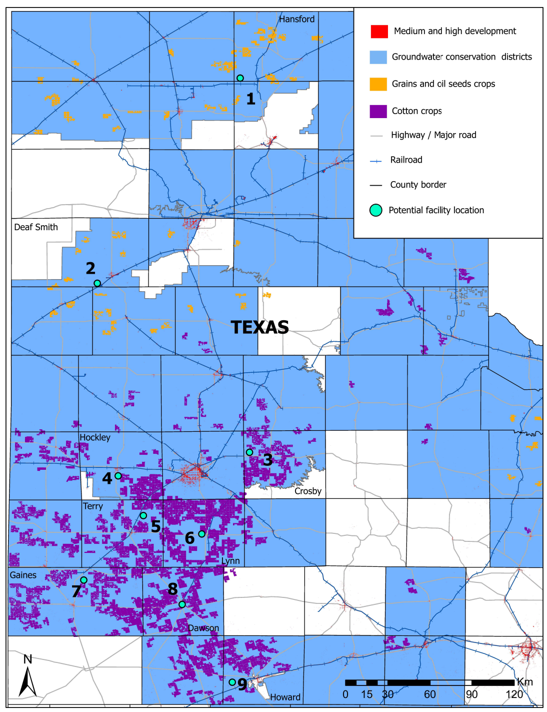


Figure 2.13: Northwestern Texas' GIS information

Chapter 3

# Optimal Production Planning and Machinery Scheduling for Semi-Arid Farms

This research proposes an integer linear optimization (ILO) model to identify optimal decisions for production planning and machinery scheduling of perennial and non-perennial crops that maximize the farmers' NPV. Key differences between this paper and previous work include the integration of both, multi-crop planning and multi-machinery scheduling into a single model while enabling crop rotation within a defined set of time periods. In contrast with the available literature, this paper differentiates multi-crop production between perennial and non-perennial crops throughout a planning horizon (e.g. a five or ten years planning). Additionally, the optimal multi-machinery scheduling is identified for each time period based on the perennial or non-perennial crop requirements. The revenue is determined based on the sales of each crop and the total costs include the machinery loan amortization, operational costs (labor, repairs, maintenance, fuel, and lube), machinery transportation costs, fertilizers, herbicides, insecticides, seeds, and irrigation water costs. For scheduling purposes, the optimal routes between the parcels are determined based on ArcGIS Pro tools. Irrigation water requirements and availability are analyzed through GIS irrigation district layers. The resulting model yields a complex large-scale ILO problem. To assess the model, two case studies for low-water-use crops are developed, one for the guar crop considering parcels located in Cochran and Hockley counties, Texas, and one for the guayule crop considering parcels located in Pinal County, Arizona. For full access to the model and input parameters, refer to [2].

## 3.1 Model Structure

One of the goals of production planning is to identify the crops to be planted in each parcel to achieve the maximum economic benefit. Likewise, one of the goals of machinery planting scheduling is to determine the sequence of parcels to work that minimizes the machinery transportation costs. The structure of the proposed model is presented in Fig. 3.1 and Fig. 3.2. Here, the optimized production planning is performed for $c$ crops in $p$ parcels, and the optimized machinery scheduling is determined for $m$ machinery sets of each crop based on $p$ parcels.

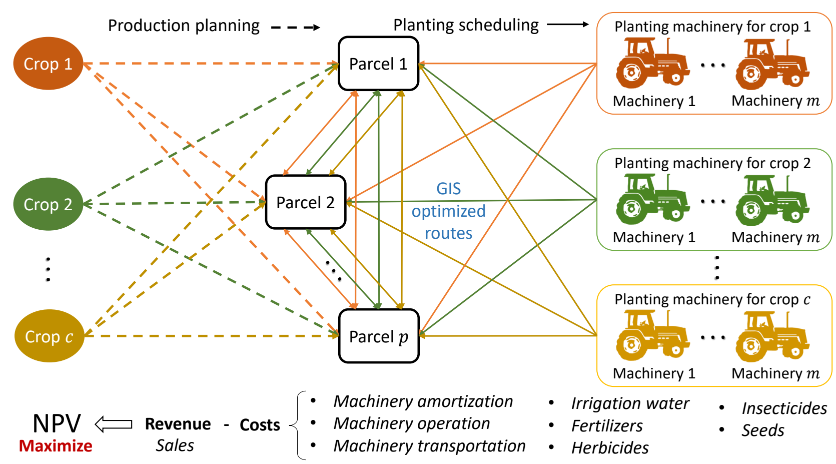


Figure 3.1: Production planning and multi-machinery scheduling: model structure (part 1)

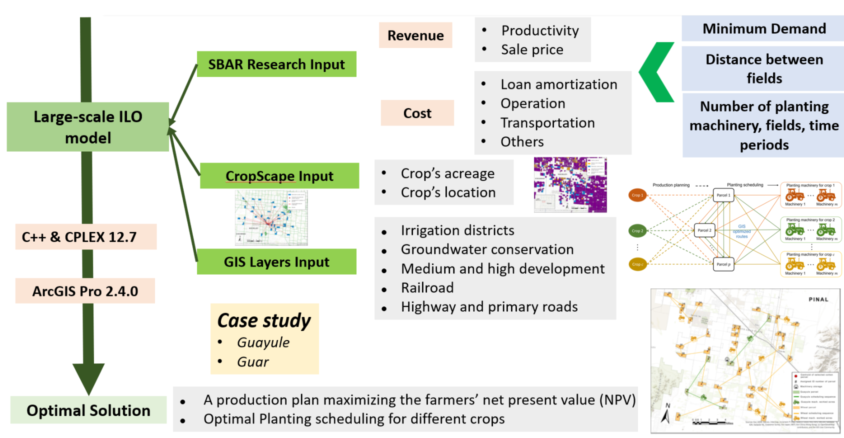


Figure 3.2: Production planning and multi-machinery scheduling: model structure (part 2)

## 3.2 GIS Data and Geoprocessing Tools

Please refer to 2.2 for detailed instructions on how to import, process, an analyze the required GIS data.

## 3.3 Defined Parameters

The sets and indices are presented in Table 3.1 and Table 3.2 defines economic, planting, and water parameters required for the model. The abbreviations of the units are defined as follows: acre (ac), $CO_2$ equivalent ($CO_2$ eq), cubic meter ($m^3$), hour (hr), kilometer (km), U.S. dollar ($), million U.S. dollar (M$), metric tonne (t), and thousand metric tonne (Kt).

Table 3.1: Optimal Production Planning and Scheduling: Sets and Indices

| | |
|---|---|
| $\mathcal{C}$: | Set of crops, indexed by $c$. |
| $\mathcal{I}$: | Set of irrigation districts, indexed by $i$. |
| $\mathcal{M}$: | Set of planting machinery, indexed by $m$, $M = \lvert\mathcal{M}\rvert$. |
| $\mathcal{P}$: | Set of parcels, indexed by $p$ and $q$, $P = \lvert\mathcal{P}\rvert$. |
| $\mathcal{R}$: | Set of route distances between parcels, indexed by $r$. |
| $\mathcal{P}_i$: | Set of parcels in irrigation district $i$. |
| $\mathcal{T}$: | Set of time periods, indexed by $t$, $T = \lvert\mathcal{T}\rvert$. |

Table 3.2: Optimal Production Planning and Scheduling: Input Parameters

| | |
|---|---|
| $C_c^{A,t}$: | Loan amortization of machinery required for planting crop $c$ for time $t$ ($). |
| $C_c^{O}$: | Machinery operational cost (including labor) during planting of crop $c$ ($/acre). |
| $C_c^{fhi}$: | Fertilizer, herbicide, and insecticide cost of crop $c$, e.g. $N_2$, P, prowl ($/acre). |
| $C_c^{T}$: | Transportation cost of planting machinery for crop $c$ ($/mile). |
| $C^{I}$: | Irrigation water cost ($/$m^3$). |
| $C^{IL}$: | Irrigation labor cost ($/acre). |
| $C_c^{S}$: | Seed cost of crop $c$ ($/acre). |
| $D_c^{t}$: | Minimum demand of crop $c$ for time period $t$ (t). |
| $S_c$: | Sale price of crop $c$ ($/t). |
| $IR^{Sale}$: | Sale price inflation rate (unitless). |
| $IR^{Cost}$: | Costs inflation rate (unitless). |
| $\alpha$: | Discount rate (unitless). |
| $T_c$: | Time periods required for crop $c$ (i.e. length of harvest season). |
| $N_c^{R}$: | Maximum number of sequential plantings (harvest length season) of crop $c$ before crop rotation. |
| $Y_{cp}$: | Yield for crop $c$ per acre of parcel $p$ (t/acre). |
| $A_{cp}$: | Acres for crop $c$ of parcel $p$ (acre). |
| $P_{cp}$: | Productivity for crop $c$ of parcel $p$ , e.g. $P_{cp} = Y_{cp} \times A_p$ (t). |
| $D_(p,q)$: | Distance $d$ between (p,q) parcels (miles). |
| $M^{Cap}$: | Capacity of planting machinery per length of harvest season (acres). |
| $MaxP$: | Maximum number of parcels the planting machinery can work on. |
| $W_c^{t}$: | Irrigation water use per acre for crop $c$ in time period $t$ ($m^3$/acre). |
| $W_{cp}^{F}$: | Maximum irrigation water capacity for crop of parcel $p$ ($m^3$). |
| $W_i^{I}$: | Water availability in irrigation district $i$ ($m^3$). |

### 3.3.1 CPLEX Parameter File Creation

A Microsoft Excel Macro-Enabled file is provided to created the input parameter file required for the CPLEX optimization software (refer to the file Product-Plan_Data_Original.xlsm) Fig. 3.3 to Fig. 3.5 will guide you through the four steps in the xlsm file. Fig. 3.3 highlights the instruction for the 4 steps, where step 1 is just to provide the desired name for the output TXT file, which will be used as an input for CPLEX. Step 1 additionally requires to fill the cardinality of the sets defined in Table 3.1, e.g., the number of crops, number of irrigation districts, and number of time periods. Once all the information in Step 1 has been completed, Step 2 just requires to click the Restructure Table button and the spreadsheet will restructure based on the provided cardinality of the sets as shown in Fig. 3.4.

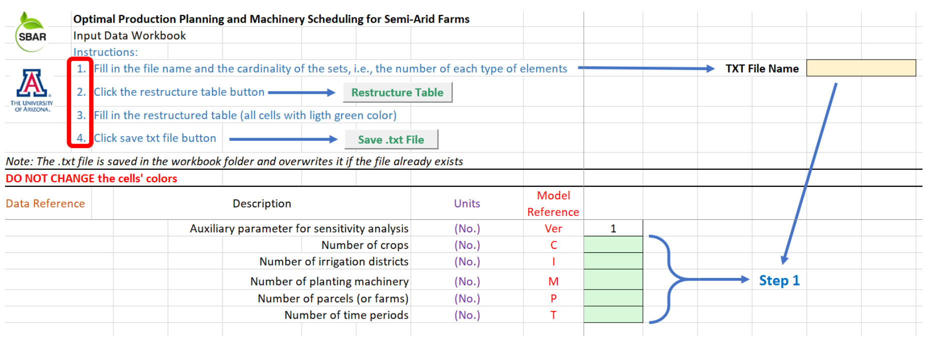


Figure 3.3: Input parameter walkthrough: Step 1

Once all of the information of the restructured table has been completed, i.e., every cell in green color as shown in Fig. 3.5, Step 4 is to click the Save .txt File button and a text file with the assigned name will be created within the same folder of the .xlsm file.

Note: **Do not use the original .xlsm file**. Since the file uses macros for the restructuring, changes cannot be easily undone. It is better to create a copy of the .xlsm file to work in and leave the original as a backup. Additionally, **the color of the cells must not be modified** since the macros are color coded. If an update in a parameter is required (except for the sets cardinalities) he values can be changed directly and the filed saved again with the Save .txt File button to update the text file.

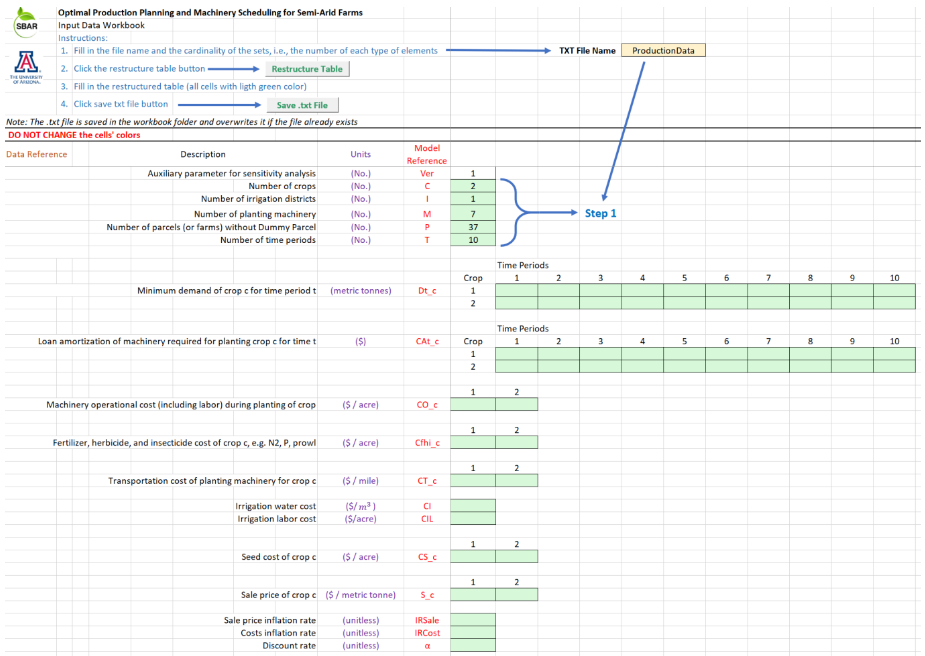


Figure 3.4: Input parameter walkthrough: Step 2

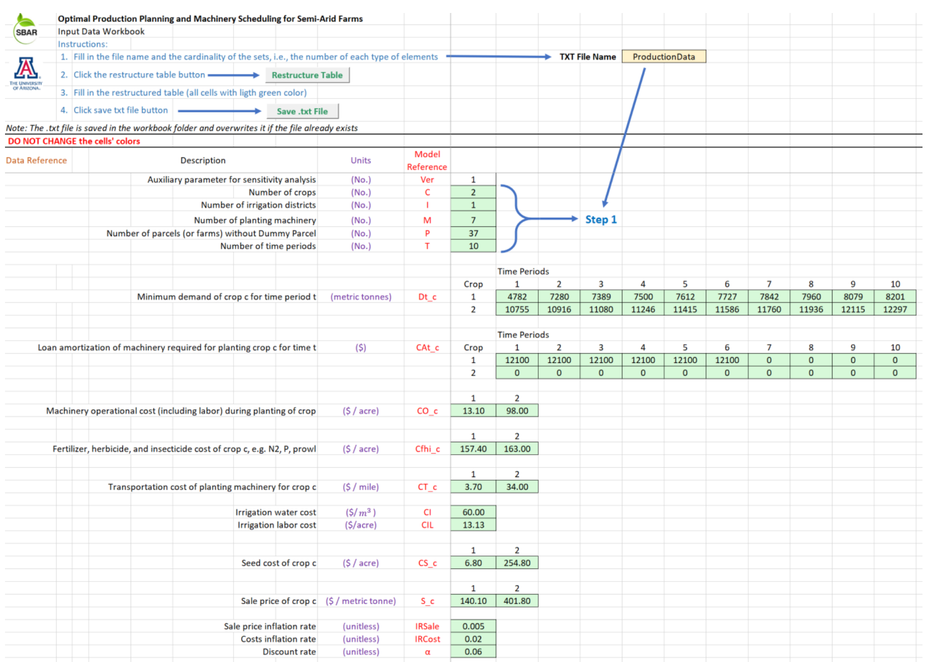


Figure 3.5: Input parameter walkthrough: Steps 3 and 4

## 3.4 Optimization Tools and Methods

Please refer to 2.4 for detailed instructions on how to access the CPLEX optimization solver through the HPC.

### 3.4.1 Reporting Outputs

Table 3.3: Optimal Production Planning and Scheduling: Decision Variables

| | |
|---|---|
| $u_{mc}$: | Binary variable that is 1 if planting machinery $m$ for crop $c$ is added and 0 otherwise ($\boldsymbol{u}$ is the vector form). |
| $w_p$: | Integer variables that denote the parcel $p$ position on the planting machinery tour, ($\boldsymbol{w}$ is the vector form). |
| $x_{cp}^t$: | Binary variable that is 1 if crop $c$ is planted in parcel $p$ for time period $t$ and 0 otherwise ($\boldsymbol{x}$ is the vector form). |
| $y_{(p,q)c}^t$: | Binary variable that is 1 if planting machinery has been assigned for crop $c$ from parcel $p$ to parcel $q$ in time period $t$ and 0 otherwise ($\boldsymbol{y}$ is the vector form). |

## 3.5 Model Assumptions

Model assumptions are specified as follows: i) the length of harvest season must be equal or greater than two time periods, $T_c \geq 2$. For example, if the length of the harvest season of a crop if one year, it needs to be modeled as two semesters; ii) the model assumes a single machinery storage location, i.e., all the planting machinery must depart from the same storage location and return to it after its work is finished; and iii) if the acreage of a parcel exceeds the planting capacity of the machinery $M^{Cap}$, the parcel needs to be divided into multiple parcels to ensure the planting machinery capacity is greater.

**Chapter 4**

# Demand-Driven Harvest Planning and Machinery Scheduling

This research proposes a large-scale mixed-integer linear optimization (MILO) model to identify optimal decisions for harvest planning and machinery scheduling that minimize the total costs. This work is an extension of previous research, considering many factors such as periodic harvest, different demands, time-window qualification, multi-machinery scheduling, resource limitations and late penalties to obtain an optimal long-term harvest plan. The goal of this optimization model is to address three significant problems when making a harvest plan: When to harvest? How to harvest? How much to harvest? To address these problems, the long-term harvest plan is decomposed into several harvest seasons. And in each harvest season, time-window variables are used to indicate the best harvest time and decision-making variables are also introduced to represent the optimal harvest scheduling. The objective function of this MILO model is defined as the total cost (including rent cost, operation cost, transportation cost and penalty cost), and model constraints consist of harvesting requirements, resource demands, harvesting capacity, time-window and transportation constraints. When solving the optimization model, we obtain a 10-year harvest plan for the crops as well as the optimal harvest scheduling during each harvest season. To validate the model, 37 fields located in Pinal County, Arizona are developed for the numerical experiments. For full access to the model and input parameters, please refer to [3].

## 4.1 Model Structure

To model the harvest planning and machinery scheduling, several inputs and geographical information are required, as shown in Fig. 4.1. The research input including various costs, diverse characteristics of crops, different harvesting demands is incorporated in this MILO model. Most of the data comes from previous research [1, 2] and some parameters are determined by researchers and industry experts from SBAR project [8]. The proposed optimization model is implemented in Python 3.9.6 using Gurobi 9.5.0 with default settings. And ArcGIS Pro 2.9.2 is used to present the GIS information and results of optimal machinery scheduling.

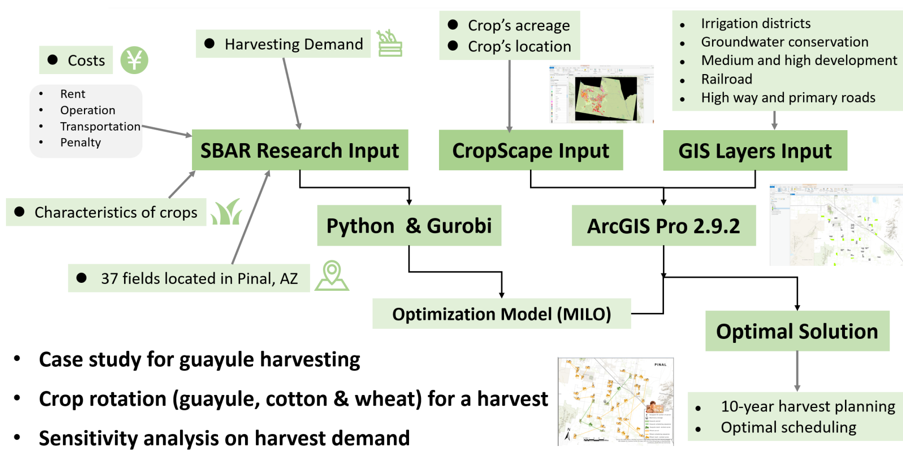


Figure 4.1: Harvest planning and machinery scheduling: model structure

## 4.2 GIS Data and Geoprocessing Tools

Please refer to 2.2 for detailed instructions on how to import, process, an analyze the required GIS data.

## 4.3 Defined Parameters

The sets and indices are presented in Table 4.1 and Table 4.2 defines several parameters required for the model.

Table 4.1: Harvest Planning and Machinery Scheduling: Sets and Indices

| | |
|---|---|
| $F$: | Set of fields, indexed by $f$ |
| $H$: | Set of harvesters, indexed by $h$ |
| $T$: | Set of time periods, indexed by $t$ (e.g. 1st week of harvesting) |
| $K$: | Set of harvest seasons, indexed by $k$ (e.g. fall of the second year) |

Table 4.2: Harvest Planning and Machinery Scheduling: Input Parameters

| | |
|---|---|
| $C_h$: | Ownership/Rent cost of harvester $h$ ($/harvest season) |
| $C^o$: | Unit machinery operational cost (including labor, repairs, maintenance, fuel, and lube) during harvesting one field ($/acre) |
| $C^r$: | Unit route/transportation cost ($/mile) |
| $d_{(i,j)}$: | Distance between field $i$ and $j$ (mile) |
| $C_f^p$: | Penalty cost incurred for not harvesting field $f$ on time ($/(acre*period)) |
| $R_f^k$: | Ready date of field $f$ for harvest season $k$ |
| $D_f^k$: | Due date of field $f$ for harvest season $k$ |
| $A_f$: | Acres for field $f$ (acre) |
| $M_f$: | Maximum number of machines that can harvest field $f$ in a given time period |
| $M^{Cap}$: | Harvesting capacity of one harvester per length of harvest season (acre) |
| $I_f^k$: | Binary values indicate whether field $f$ should be harvested in the harvest season $k$ |

### 4.3.1 Input Parameter File Creation

A Microsoft Excel Macro-Enabled file is provided to created the input parameter file required for the Gurobi optimization software (refer to the file HarvestPlan_data_Original.xlsm) Fig. 4.2 to Fig. 4.4 will guide you through the four steps in the xlsm file. Fig. 4.2 highlights the instruction for the 4 steps, where step 1 is just to provide the desired name for the output TXT file, which will be used as an input for Gurobi. Step 1 additionally requires to fill the cardinality of the sets defined in Table 4.1, e.g., the number of fields, number of harvesters, number of harvest seasons, and number of time periods. Once all the information in Step 1 has been completed, Step 2 just requires to click the Restructure Table button and the spreadsheet will restructure based on the provided cardinality of the sets as shown in Fig. 4.3.

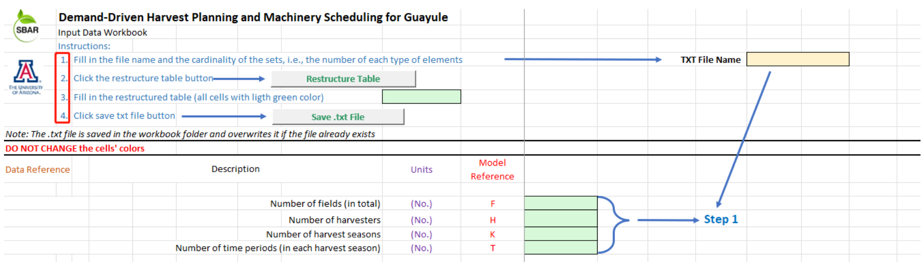


Figure 4.2: Input parameter walkthrough: Step 1

Once all of the information of the restructured table has been completed, i.e., every cell in green color as shown in Fig. 4.4, Step 4 is to click the Save .txt File button and a text file with the assigned name will be created within the same folder of the .xlsm file.

Note: **Do not use the original .xlsm file**. Since the file uses macros for the restructuring, changes cannot be easily undone. It is better to create a copy of the .xlsm file to work in and leave the original as a backup. Additionally, **the color of the cells must not be modified** since the macros are color coded. If an update in a parameter is required, the values can be changed directly and the filed saved again with the Save .txt File button to update the text file.

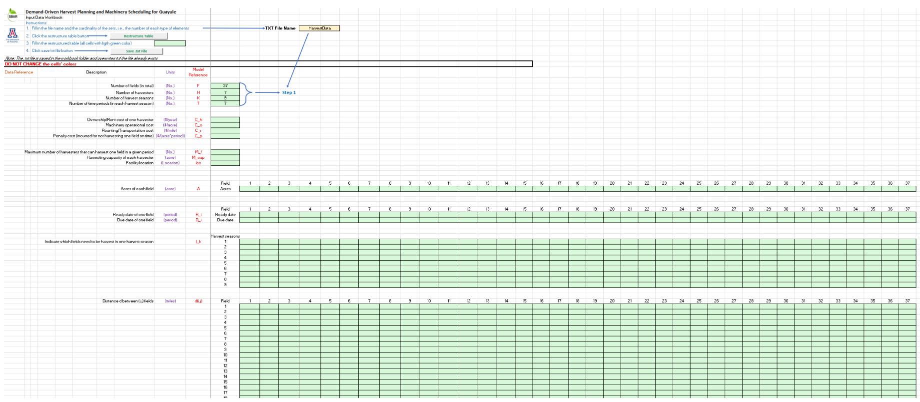

Figure 4.3: Input parameter walkthrough: Step 2

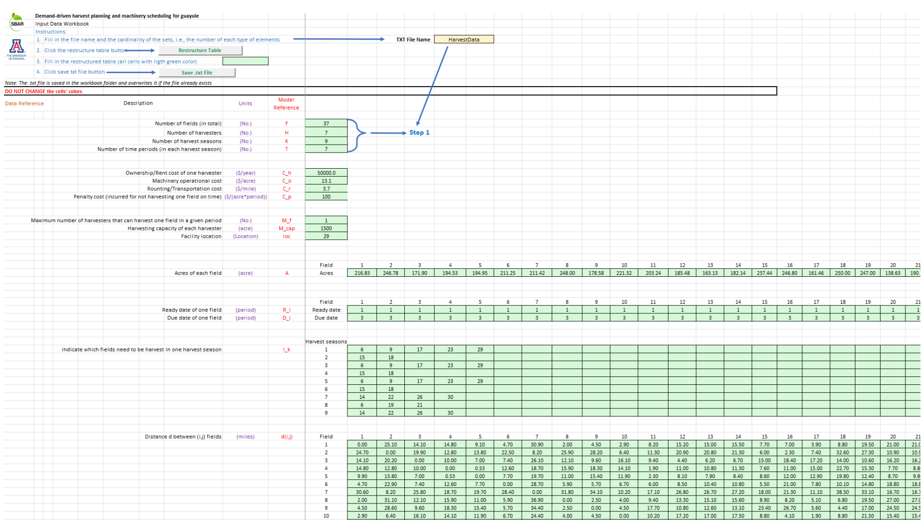

Figure 4.4: Input parameter walkthrough: Steps 3 and 4

## 4.4 Optimization Tools and Methods

### 4.4.1 Gurobi Optimizer Through a High Performance Computer

A simple way to access the Gurobi optimization solver through the HPC is using the Open OnDemand website (https://ood.hpc.arizona.edu/), which is an NSF-funded open-source HPC portal and is available for users and provides web browser access for interfacing with HPC. Fig. 4.5 to Fig. 4.7 will guide you through the three steps to run Python with Gurobi optimizer through HPC.

After entering the OnDemand website, go to "My Interactive Sessions" → "Jupyter Notebook". Once you've entered all your details, click "Launch" at the bottom of the page. This will create a job request for Jupyter Notebook in the HPC. After a few minutes (waiting time depends on the parameters you entered), you can click "Connect to Jupyter" to start Jupyter Notebook in the HPC as shown in Fig. 4.6. After that, you can use Jupyter Notebook to view, edit, and upload/download your files. In the Jupyter Notebook, find the corresponding ipynb file and click "Run All" under the "Cell" button. Then the optimization model will be implemented by Python with Gurobi optimizer through HPC (see Fig. 4.7).

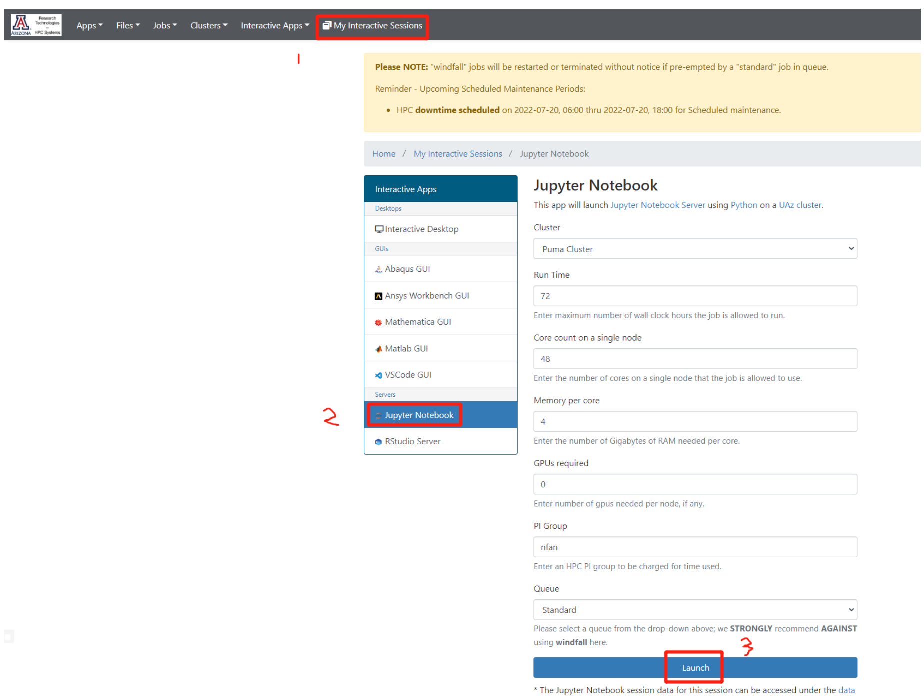


Figure 4.5: Running Gurobi optimizer in the HPC (Part 1)

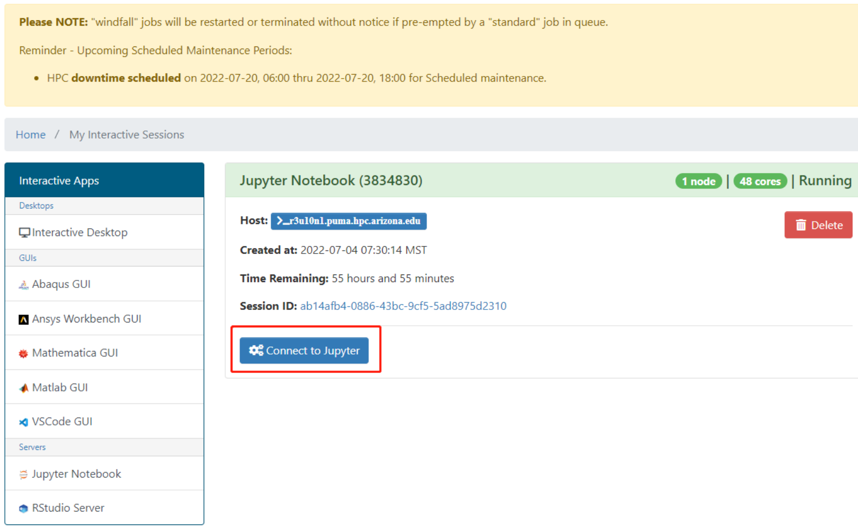


Figure 4.6: Running Gurobi optimizer in the HPC (Part 2)

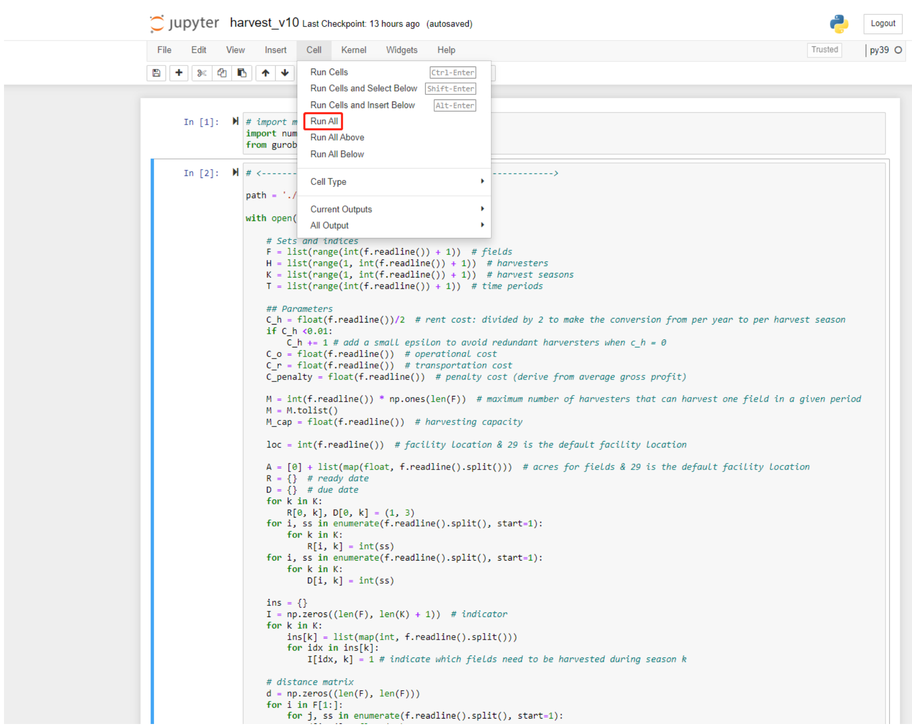


Figure 4.7: Running Gurobi optimizer in the HPC (Part 3)

### 4.4.2 Reporting Outputs

Table 4.3: Harvest Planning and Machinery Scheduling: Decision Variables

| | |
|---|---|
| $u_h^k \in \{0,1\}$: | $u_h^k = 1$ if harvester $h$ is used in harvest season $k$ |
| $x_{ft}^k \in \{0,1\}$: | $x_{ft}^k = 1$ if field $f$ is harvested at period $t$ in harvest season $k$ |
| $y_{ht,(i,j)}^k \in \{0,1\}$: | $y_{ht,(i,j)}^k = 1$ if harvester $h$ moves from field $i$ to $j$ at the end of period $t$ of harvest season $k$ |
| $z_{hft}^k \in \{0,1\}$: | $z_{hft}^k = 1$ if harvester $h$ for field $f$ is used at period $t$ in harvest season $k$ |
| $\alpha_{hft}^k \in [0,1]$: | Harvesting fraction for field $f$ using harvester $h$ in period $t$ of harvest season $k$ |
| $s_f^k \in \{0,\ldots,\lvert T\rvert\}$: | Harvest starting period for field $f$ in harvest season $k$ |
| $e_f^k \in \{0,\ldots,\lvert T\rvert\}$: | Harvest ending period for field $f$ in harvest season $k$ |

## 4.5 Model Assumptions

Model assumptions are specified as follows: 1) Suppose a time period is one week, and guayule is harvested in each field in a time window $[R_f^k, D_f^k]$ during the harvest season $k$. 2) Once a field is available, it should be harvested before the due date. Otherwise, there may exist some penalties for not harvesting on time. 3) Guayule in all fields should be completely harvested over the harvest season. 4) Suppose there exists only one warehouse/facility in this harvest area, and all machinery departs from the warehouse/facility at the beginning of a harvest season and returns to it at the end of the harvest season. 5) Machinery moves between fields at the end of the period in order to fully utilize productive time during the next working period.

**Chapter 5**

# Conclusions

This instruction manual summarizes the development and current capabilities of the BSCT model. The instruction manual focuses on three major sections. One on a a stochastic optimization model to achieve optimal decisions for facility locations, transportations from field to facilities, and finally to customers, with a multi-objective approach to quantify the economic benefits, environmental, and social impacts. This model is developed based on the adoption requirements of the two desert-dwelling crops guar and guayule and modifications may be required to assess the adoption of other crops. To capture the uncertainty of the adoption rates from each farm, stochastic scenarios were applied. The model and algorithm were tested on a case study for two desert-dwelling crops, guayule in Arizona for Maricopa and Pinal counties, and guar in New Mexico for Dona Ana County (please refer to [1]).

The second major section focuses on an optimization model to achieve optimal operations for farm production planning and machinery scheduling for perennial and non-perennial crops considering multi-crop planning, multi- machinery scheduling, and crop rotation within a defined set of time periods. Two case studies are developed, one for the guar crop (non-perennial) considering parcels located in the Cochran and Hockley counties, TX, and one for the guayule crop (6-year perennial) considering parcels located in Pinal County, AZ (please refer to [2]).

The third major section focuses on an optimization model to obtain optimal harvest planning and machinery scheduling under a series of factors like periodic harvest, different demands, resource limitations, etc. In addition to satisfy different types of requirements, this optimization model also aims to find the best decision-making strategy in order to minimize the total cost. Case studies for three crops located in Pinal County, Arizona are developed to validate the model (please refer to [3]).

We consider that BSCT will provide a fundamental framework for further facility location, production and scheduling optimization research. Future studies may consider complementing our model with optimization approaches for production planning and machinery scheduling in semi-arid regions to maximize the different stakeholders' benefits. Additional research to increase the sale value of guar and guayule may be required for their adoption in the Southwestern U.S. agriculture. Rainy weather may affect the harvesting scheduling due to level of moisture requirements for the processing facilities. Therefore, further weather analysis considering rain risk during the harvesting periods can be considered in future research. Additionally, a water life-cycle-analysis per crop can be performed to identify farming regions potential water availability risks which may lead to consider higher adoption rates for guayule and guar within those regions.

# Bibliography


[1] Daniel A. Zuniga Vazquez, Ou Sun, Neng Fan, Evan Sproul, Hailey M. Summers, Jason C. Quinn, Sita Khanal, Paul Gutierrez, VeeAnder Mealing, Amy E. Landis, Clark Seavert, Trent Teegerstrom, and Blase Evancho. Integrating environmental and social impacts into optimal design of guayule and guar supply chains. *Computers & Chemical Engineering*, 146:107223, 2021.

[2] Daniel A. Zuniga Vazquez, Neng Fan, Trent Teegerstrom, Clark Seavert, Hailey M. Summers, Evan Sproul, and Jason C. Quinn. Optimal production planning and machinery scheduling for semi-arid farms. *Computers and Electronics in Agriculture*, 2021. accepted manuscript.

[3] Shunyu Yao, Neng Fan, and et al. Demand-driven harvest planning and machinery scheduling for guayule. 2022. in preparation.

[4] USDA. United states department of agriculture. irrigation & water use, 2019. https://www.ers.usda.gov/topics/farm-practices-management/irrigation-water-use (accessed December 2, 2020).

[5] Mehran Homayounfar, Sai Hin Laiand, Mehdi Zomorodian, Ali Reza Sepaskhah, and Arman Ganji. Optimal crop water allocation in case of drought occurrence, imposing deficit irrigation with proportional cutback constraint. *Water Resources Management*, 28:3207 – 3225, 2014.

[6] G. Niu, Y.P. Li, G.H. Huang, J. Liu, and Y.R. Fan. Crop planning and water resource allocation for sustainable development of an irrigation region in china under multiple uncertainties. *Agricultural Water Management*, 166:53 – 69, 2016.

[7] Chenglong Zhang, Mo Li, and Ping Guo. An interval multistage joint-probabilistic chance-constrained programming model with left-hand-side randomness for crop area planning under uncertainty. *Journal of Cleaner Production*, 167:1276 – 1289, 2017.

[8] SBAR. Sustainable bioeconomy for arid regions. project objectives, 2017. https://sbar.arizona.edu/our-goals/project-objectives (accessed July 21, 2020).

[9] Griffin M. Bates and Katrina Cornish. Rapid and complete removal of guayule (parthenium argentatum) leaves by cryodefoliation, and freeze and thaw induction of rubber particle coagulation. *Industrial Crops and Products*, 125:491 – 495, 2018.

[10] Daniel C. Ilut, Paul L. Sanchez, Terry A. Coffelt, John M. Dyer, Matthew A. Jenks, and Michael A. Gore. A century of guayule: Comprehensive genetic characterization of the guayule (parthenium argentatum a. gray) usda germplasm collection. *bioRxiv*, 2017.

[11] Abdulraheim Hasan and Manar El-Sayed Abdel-Raouf. Applications of guar gum and its derivatives in petroleum industry: A review. *Egyptian Journal of Petroleum*, 27, 05 2018.

[12] Archana George, Priyanka A. Shah, and Pranav S. Shrivastav. Guar gum: Versatile natural polymer for drug delivery applications. *European Polymer Journal*, 112:722 – 735, 2019.

[13] USDA. United states department of agriculture. census of agriculture, 2018. https://www.nass.usda.gov/AgCensus/ (accessed January 28, 2020).

[14] NASS. National agricultural statistics service. cropscape - cropland data layer, 2019. https://nassgeodata.gmu.edu/CropScape/ (accessed January 28, 2020).