Development of Complementary Fresh-Food Systems Through the Exploration and Identification of Profit-Maximizing, Supply Chains

by

Hector M. Flores

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved October 2017 by the
Graduate Supervisory Committee:

Rene Villalobos, Chair
Rong Pan
Teresa Wu
Nathan Parker

ARIZONA STATE UNIVERSITY

December 2017

ABSTRACT

One of the greatest 21st century challenges is meeting the needs of a growing world population expected to increase 35% by 2050 given projected trends in diets, consumption and income. This in turn requires a 70-100% improvement on current production capability, even as the world is undergoing systemic climate pattern changes. This growth not only translates to higher demand for staple products, such as rice, wheat, and beans, but also creates demand for high-value products such as fresh fruits and vegetables (FVs), fueled by better economic conditions and a more health conscious consumer. In this case, it would seem that these trends would present opportunities for the economic development of environmentally well-suited regions to produce high-value products. Interestingly, many regions with production potential still exhibit a considerable gap between their current and ‘true’ maximum capability, especially in places where poverty is more common. Paradoxically, often high-value, horticultural products could be produced in these regions, if relatively small capital investments are made and proper marketing and distribution channels are created. The hypothesis is that small farmers within local agricultural systems are well positioned to take advantage of existing sustainable and profitable opportunities, specifically in high-value agricultural production. Unearthing these opportunities can entice investments in small farming development and help them enter the horticultural industry, thus expand the volume, variety and/or quality of products available for global consumption. In this dissertation, the objective is three-fold: (1) to demonstrate the hidden production potential that exist within local agricultural communities, (2) highlight the importance of supply chain modeling tools in the strategic

design of local agricultural systems, and (3) demonstrate the application of optimization and machine learning techniques to strategize the implementation of protective agricultural technologies.

As part of this dissertation, a yield approximation method is developed and integrated with a mixed-integer program to estimate a region's potential to produce non-perennial, vegetable items. This integration offers practical approximations that help decision-makers identify technologies needed to protect agricultural production, alter harvesting patterns to better match market behavior, and provide an analytical framework through which external investment entities can assess different production options

TABLE OF CONTENTS

CHAPTER Page

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

1. INTRODUCTION

One of the greatest 21st century challenges is meeting the needs of a growing world population expected to increase 35% by 2050 given projected trends in diets, consumption and income. This in turn requires a 70-100% improvement on current production capability (van Wart et al., 2013). This growth in world population not only translates to higher demand for staple products, such as rice, wheat, and beans, but also creates demand for high-value products such as fresh fruits and vegetables (FVs) fueled by better economic conditions and a more health conscious consumer. In this dynamic, it would seem that these trends would present opportunities for regions environmentally, well-suited to grow these products based on their geographical location and climate patterns. Interestingly, many of these regions (often in impoverished areas) exhibit a considerable gap between their current and maximum production capability attained under better agricultural practices and/or technologies. Interestingly, high demand/value agricultural products could be produced in these regions, if relatively small capital investments are made in technology and/or knowledge, and appropriate marketing and distribution channels are made available. The work presented in this dissertation tackles this problem from the perspective of the design and planning of revenue-maximizing, agri-food supply chains. Specifically, focus is given to fresh fruits and vegetables supply chains, since they have the potential to significantly increase the income of producing regions.

The base hypothesis of this dissertation is that there exist unidentified, geographical regions with adequate climate patterns to grow high-value agricultural products in a profitable and sustainable manner. The ability to identify and assess these opportunities can entice micro and small farmers to participate in agricultural supply chains, as well as incentivize external investments into new production implementations. Thus, the development of agricultural planning models that can incorporate rough, production estimates of high-value plants based on a region's temporal environmental conditions is an important first step. This dissertation presents a methodology that approximates the expected yield of non-perennial, FVs as function of basic environmental parameters, such as maximum and minimum temperatures, observed under controlled, production settings. These functions can then be used to approximate expected FV production in areas where availability of environmental data is far more limited or can only be estimated. As part of this process, empirical information can be used to adjust the pattern structure of these yield functions as new information is acquired. In this case, we note that although these approximation methods are less exact than more sophisticated plant growth models, it offers practical approximations that may help decision-makers identify technologies needed to protect agricultural production, alter harvesting patterns to better match market behavior, and provide an analytical framework through which external investment entities can assess different production opportunities.

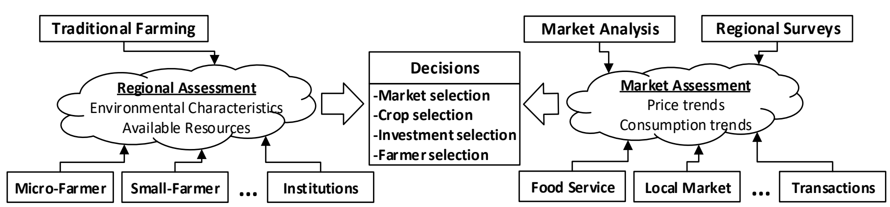


Figure 1-1: Framework for the Design of a Local Fresh-Food System

The success of a fresh food system often depends on the variety and length of the product offering, which can often be achieved by either expanding current operations or through partnerships with other food systems that can complement production. In this work, a supply chain design is constructed based on coordinated production from a partnership of local fresh food systems in regions with complementary seasonal weather patterns (Figure 1-1). The hypothesis is that small farmers within each system can coordinate their production to circumvent limited production windows and expand the variety and length of their offering. Furthermore, given price volatility and seasonality of product availability within U.S. wholesale markets, consumption points external to the immediate local system are included in this design as they have large demand potential, are easier to access, and often exhibit temporal price opportunities that can enhance profitability through market-to-market shipments (Flores and Villalobos, 2013). Figure 1-2 presents the basic set-up for this system, in which each circle represents a local fresh-food system with different climate patterns. In this representation, systems 2 and 3 have sufficient internal consumption market to serve, while system 1 only serves the external markets. These systems consolidate their production in a distribution center and sell to external markets based on their observed price patterns. The assessment framework then seeks to construct a harvesting schedule that matches the temperature and price behavior of the regions and markets, respectively, to maximize the profitability of a centralized decision-maker. As part of the optimization framework, an optimal logistic configuration is developed that moves the product from its production source to the end consumer market at a more strategic level.

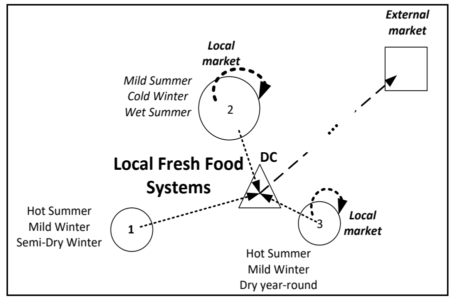


Figure 1-2: Design of a Local Fresh-Food System

The first objective of this dissertation is to develop a deterministic assessment framework that maximizes the profitability of a centralized decision-maker that attempts to identify and coordinate a partnership of local fresh-food system and complementary production. The output from this framework is a planting and harvesting schedule for each of the considered regions and their assigned consumer market. This would allow investors, small farmers, or potential entrants to assess the profitability of specific product, region, and market configurations given different production technologies. The inputs to this framework include historic information on consumer market prices for different vegetable items and environmental data of select, geographic regions. It is worth noting that the initial assessment would take the perspective of a centralized decision-maker in identifying optimal decisions that maximizes the profitability of the operations as a whole. The second objective is to consider the stochasticity of the different parameters in the deterministic formulation. The main task will be the development of techniques that can incorporate a larger set of potential scenarios in environmental and market conditions within different fresh-food systems. The final objective in this dissertation is the decentralization of the initial formulation to consider each of the participating farmers as individual entities with their own minimum profit requirements in newly producing systems. Nonetheless, the overall encapsulating goal is to improve the odds of implementing a successful, local fresh-food operation, which in turn encourages new participants into the agricultural supply chain, and thereby, expands the volume, variety and quality of products available for local and global consumption.

1.1 Problem Definition

Closing the difference between projected food consumption and future production capability, also known as the yield gap, has been an important area of research for agricultural researchers over the past two decades. There is a wide consensus that the yield gap can be reduced through improved technologies, better training to farmers, and increased resources in under-producing regions. However, incentivizing investments in under producing regions has been difficult, since many of these regions produce staple products whose market profitability is not high enough to attract potential partnerships from investors. Furthermore, little to no research exists that aims to identify regions with the highest potential to produce alternative, high-value products, which could incentivize such investments. Moreover, once a region has been identified, little amount of research exists that details the specific methods through which farmers can be incorporated into the horticultural supply chains. This is important because very often the investments made by governmental entities on physical infrastructure do not give the expected results because farmers and the underlying supply chain conditions are not taken into consideration.

To address this problem, the problem is defined from the perspective of a single investing entity assessing the capability of geographical regions to produce horticultural products that have shown profitability in the market place, which is comparable to other financial, investment instruments. One should note that this entity may or may not be related to the agricultural community but may be seeking to invest in the agricultural development of a region. In this case, it is assumed that the entity has access to adequate information technologies that would allow him to observe, analyze, and identify market opportunities, as well as the ability to raise enough capital to invest in the technological upgrades of the identified regions. It would also be assumed that this entity would have access to market, as well as proper transportation equipment that would allow him to obtain the products from the identified region to transitory distribution centers and onto the market place. Finally, we also consider the farmers within the marginalized regions to constitute separate entities with their own set of incentives and production constraints.

The first part of this dissertation develops tools that incorporate the environmental (e.g. temperature, precipitation) and technological characteristics (e.g. greenhouse, fertilizers, etc.) of different regions into a planning model in order to determine their capability to produce high-value horticultural items. This set of items will be based on market opportunities identified by the main producer for which the famer may be constrained by his own environmental limitation to produce. In this case, the main farmer may find it profitable to integrate the production capability of another region to complement his own production, in order to capture identified market opportunities. Additionally, the introduction of additional production regions can be used as a way to protect against climate variability in his own region. The task is then to determine based on the environmental and technological constraints of each region, as well as the profitability of the opportunity, the specific technology and region on which the primary investing entity should invest in order to maximize his profits. Furthermore, the stochastic characteristics of the climate-based parameters, such as temperature and yields, will have to be considered as part of this exploration. Finally, since the geographical regions are considered independent, the modeling structure allows the use of methods that can decompose the problem into more efficiently solvable components.

Once a set of regions has been identified, the second phase of the problem is the development of tools that can optimally allocate technologies *within* an identified region. In this case, each party involved would have their own set of profit incentives and would have to decide whether to participate in the production of high-value fresh fruits and vegetables. Similarly, the entity making the investment would have to decide the allocation of the implemented technological infrastructure space (e.g. greenhouses, protective technologies), as well as the production targets for each individual farmer based on the harvest schedule provided by first phase results. Different manners to model the alignment of these conditions will be assessed in order to determine a production agreement that would meet the profit requirements of each participating party.

The research scope will be geared towards the development of mostly strategic and tactical level planning tools. Strategic level planning tools will seek to determine the set of products that could potentially be commercialized based on established operations of the primary farmer, as well as any new identified market opportunities. Based on the set of selected products, potential production regions will be identified, from which the primary farmer will select those regions to incorporate. Furthermore, since tactical level planning also drives many of the decisions at the strategic level (e.g. quantity and timing of planting/harvesting), this kind of tools will be incorporated. For purposes of this research, operational level planning will not be considered.

1.2 Contributions of the Dissertation Research

The work to be presented will be one of the first works focusing on the development of tools that seek to model and identify geographical regions with the potential to produce high-value, horticultural products. Among the contributions of this work is the design of stochastic programs that seek to match resource and environmental characteristics of geographical regions to the production requirements of horticultural items. There are several complicating aspects of this problem, such as the stochastic characteristics of several parameters involved in the model formulation, that have so far not been considered under the context of strategically sourcing agricultural products from different geographical regions. Moreover, within these problem formulations, we will consider crop yield functions, which are mostly dependent on the realization of stochastic variables (e.g. temperature, precipitation) and that directly impact expected profits.

To the best of our knowledge, little to no work has been done in the strategic sourcing of horticultural production based on both market opportunity and regional production potential. In this dissertation, we will consider this problem. As part of this study, we seek to identify geographical regions with semi adequate conditions (e.g. environment, resource) to produce particular items, while identifying the limiting constraints and potential remedies for them. Consequently, the research will develop methods through which farmers can connect market demand opportunities with agricultural regions that have the production potential to produce them.

In order to find efficient solutions to this problem, the implementation of two-stage stochastic decomposition methods will be explored. These solution methods are common in other type of capacity expansion problems in which some parameters are random. In this dissertation, similar solution methods are proposed to determine optimal investment strategies for upgrading the technological infrastructure of producing regions. However, this problem context has several aspects that differentiates it from previous applications, since it attempts to match consumer demand, environmental constraints, and production requirements in an integrated manner through the strategic allocation of technologies for perishable, horticultural products.

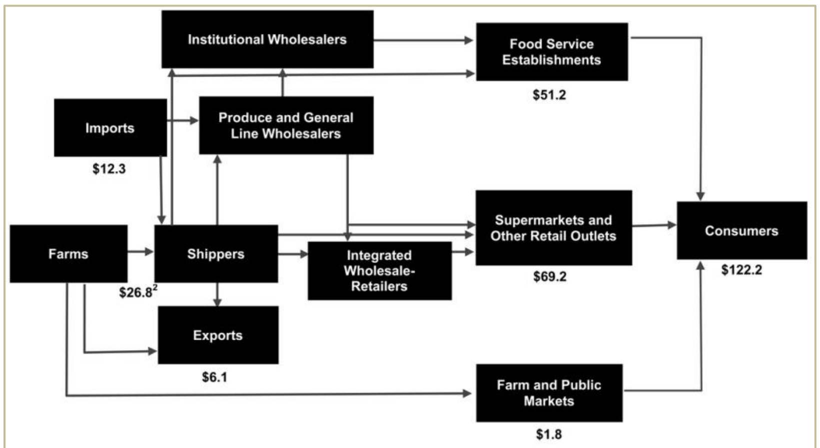
Furthermore, literature related to the strategic integration of marginalized agricultural regions into established ASCs is limited. This research effort will be among the first to develop optimization-based modeling tools that strategically source products from regions without historical production by specifically using temporal, environmental conditions and resource/logistic characteristics to identify horticultural production opportunities. Within this problem context, we would contribute optimal supply chain designs that transport products from newly identified regions to the market place, as well as optimal allocation of technologies and resources to farmers.

The work presented here will also develop models through which technologies are allocated to the different farmers in an identified region. In this case, decentralized modeling structure are proposed that consider the incentives of each farmer to transition to higher value products, which can prove to be important in shaping production behavior of agricultural farmers. The decentralized framework will also consider production and logistic decisions made within local agricultural production. Furthermore, the proposed decentralized formulation sets the framework for research expansion in the design of contract mechanisms to study the effect of different profitability profiles of different farmers and the level of participation in agricultural development.

1.3 Benefits of Research

As explained earlier, the projected increase in global consumption of agricultural products coupled with changes in climate creates the need for better ways of meeting demand with supply. The type of modeling tools as those proposed in this dissertation will be beneficial for those farmers looking for ways to capitalize on the opportunities that are bound to appear in the market place. Farmers and even policy makers will be able to use these kinds of tools to find promising under producing regions and strategically invest in them to insert them in established distribution and marketing channels. On the other hand, farmer in these sub regions will benefit from the technological investments and knowledge while upgrading their own operations. Furthermore, the ability for investors to identify production opportunities in region with under producing systems can help many farmers that are often marginalized and excluded from global markets. This can help agricultural communities increase their production capability, improve regional economies, and decrease rural poverty.

The proposed models in this research add another way of assessing a region's production potential that may otherwise be hidden by environmental and/or resource restrictions. Methods of estimating agricultural yield gap have generally focused on a region's potential to produce traditional crops, such as rice, wheat and beans, for the purpose of estimating current regional and global gaps. However, most of these assessments do not consider technologies that are deemed more sophisticated, more realistic and have the ability to push yields above the levels reported by yield potential assessments (Dietrich et al., 2014). Furthermore, the assessments have been generally geared towards land use exploration under a context of policy making. In this case, we shift the perspective to the development of tools that can be used to assess the profitability of investing in different types of regions in order to better position the small farmer. This can be assessed from the perspective of both marginalized regions, as well as local food systems.



[1] Excludes nuts and pulses, and frozen or canned fruits and vegetables.
[2] This value is larger than the farm gate value reported in Table 1 since it includes a supplemental estimated value for fresh-market fruit and vegetable production not captured by NASS/USDA.

Figure 1-3: U.S. Fresh F&Vs Value Chain (Estimated Dollar Sales) Billions

The potential profitability of being able to directly connect farmers with end consumers is substantial. For example, the U.S. fresh fruit and vegetable industry is estimated to be worth over $122.1 billion, from which farmers receive approximately $21.8 billion, according to 2010 estimates by Cook (2011) (Figure 1-3). This means that for every dollar spent by the end consumer on fruits and vegetables, the farmer receives less than $0.20, while the rest is divided among supply chain intermediaries. In this case, in order for the farmer to recoup a portion of these profit margins, the farmer would need to offer the end consumer some of the services currently provided by intermediaries; chief among them is year-round availability of products. The modeling tools developed in this dissertation research addresses this need from the perspective of the farmer as it seeks for improved commercial position in the industry.

1.4 Dissertation Overview

In the following chapters, a brief description is given into current literature related to production planning, land use exploration, integrated assessment models, and profit allocation mechanisms in agricultural related environments (Chapter 2). Then in Chapter 3, the set of proposed models are presented along with expected outcomes. In Chapter 4, a centralized, deterministic formulation of the agricultural production exploration is developed and presented along with a case-study applied to the U.S. southwest region of Arizona and New Mexico. Chapter 5 expands on the initial deterministic formulation to incorporate the stochastic components of environmental and market parameters using two solving mechanisms. Chapter 6 presents a decentralized version of the deterministic formulation in which each of the farmer entities are assumed to have their own profitability requirements applied to a small case study instance within the same U.S. southwest region. Finally, Chapter 7 presents a final dissertation discussion and areas of future work.

2. LITERATURE REVIEW

There are three main areas of research that were identified that relate problem addressed in this study. The first one deals with agricultural production planning. Since a major component to our problem is the identification of geographical regions with production potential and their incorporation into established supply chains, we searched for literature related to this problem within two main areas (1) agricultural production planning and (2) crop selection. Given that there are various sources of variability in the agricultural industry; research geared towards the development of improved planning models under variable conditions has garnered increased attention over the last few decades. However, most of the academic work in this area has been concentrated on production planning and the development of efficient supply chains of already established production sites, distribution centers, and the underlying logistics systems. As it will be noted later as part of the review, the amount of research dedicated to the exploration and identification of areas with production potential based on geographical-based environmental characteristics and profitability is limited.

The second aspect of this problem is estimating the actual production potential of an agricultural region, including undiscovered potential to produce high-value, horticultural products. For this purpose, we target academic works focused on assessment methodologies geared towards estimating the production potential of particular regions. In this case, we identified three main categories of literature work related to this problem: (1) land use exploration and management, (2) yield gap estimation, and (3) climate impact on land use. In the first area, we found works related to models aimed at determining the best way to make use of land and geographical resources. The second area involves literature related to the development of methodologies capable of providing consistent estimations of a region's yield gap. This task can prove to be difficult given regional disparities in production and geographical information. In the third area of focus, we identify works that go beyond land use exploration and yield gap estimation to that attempt to estimate future land use patterns. Research in this area has received relatively good coverage during the last two decades especially as climate change has started to impact agricultural production systems. However, most of the tools that have been developed in this area have been geared towards policy-based assessments. To the best of our knowledge, there is still a large literature gap in addressing the problem from the perspective of the individual farmer and the role of targeted technological investments to uncover new production potential.

The third aspect of this problem is the allocation of technologies/resources to farmers in marginalized regions. In this third case, we divided our focus into two main areas: (1) the allocation of technologies and resources in agricultural settings and (2) the sharing of profits and vertical cooperation among agents in traditional and agricultural supply chains. Since agricultural supply chains deal with high amount of variability, we searched for literature work dealing with revenue sharing within the supply chain under uncertain conditions. In this case, related works were found for application in bio-renewable supply chains. There also has been extensive case study work related to capacity sharing within a more traditional manufacturing setting. As it will be explained in detail, there also has been a renewed interest in contract farming, in which large corporations engage in forward contracts with farmers, while supplying the needed inputs in exchange for a quality controlled product available in a time manner for the purpose of ensuring quantity and price stability (Huh and Lall, 2013). The type of problems that have been addressed related to this topic is of relevance to

Overall, this literature review is not meant to be exhaustive, but rather it looks to point out the general direction of current research related to this problem, as well as identifying gaps in literature. This section is divided according to the three aspects of the problems described above. First, literature related to agricultural supply chains will be discussed. Secondly, academic works related to integrated assessment models will presented. Thirdly, the review of literature concludes with works related to profit allocation and revenue sharing policies. Finally, an assessment of the existing gap in literature is presented.

2.1 Agricultural Production Planning

Given the increase in the complexity and variability of the agricultural industry, especially in fresh produce, there has been a rise in the number of works geared towards the development of improved production planning agricultural supply chains. In this case, most studies develop deterministic and stochastic models to make tactical and operational decisions, such as when and how much to plant/harvest during a particular production season. An area that has not yet received much focus is the selection of crops under varying environmental and resource conditions.

2.1.1 Agricultural Production Planning

Before we begin to give a review of work in the area of agricultural production planning it is important to begin with the work of Glen (1987), who provides an extensive review of previous works pertaining to application of operations research to agricultural planning, particularly at the farmer level. Most of the works pre-1987 cited in the review use a variety of formulations of deterministic and stochastic linear, quadratic, and mixed integer programs, as well as simulation models, to make tactical and operational decisions such as machinery selection, and planting/harvesting methods. Lowe and Preckel (2004) add to this review by incorporating works done after this initial review, and although, it is not as extensive, it provides possible directions for future research in this field. Lucas and Chhajed (2004) add to this review by presenting some interesting applications of optimization models to agricultural-based problem settings. For a more complete review of agricultural planning models pertaining to agricultural supply chains the reader is referred to Ahumada and Villalobos (2009).

One thing to note in this collection of works is that most deal with tactical and operational decision-making of already established operations, that is for crops that have already attained a level of success to insure their economic sustainability. For example, Ahumada et al. (2012) propose a stochastic model for tactical level planning of the production and distribution of perishable agricultural products. In this case, the authors construct discrete scenarios, which account for different yield and demand outcomes. Similarly, in Ahumada and Villalobos (2011), the authors propose an operational level model that would help the grower make short-term production and distribution decisions during the harvesting periods once production plans and the corresponding supply chain design have been created. Interestingly, literature relating to agricultural production planning that considers the strategic sourcing of agricultural products from different geographical regions was not found in this review.

Most literature that relates to the strategic sourcing of agricultural-related products was found in works associated to the biomass industry. In this research area, most supply chain studies investigate the infrastructure allocation problem, in which the decision variables become the number of facilities, their locations and the interactions amongst them (Lucas and Chhajed, 2004). Awudu and Zhang (2012) gives a thorough review of the biomass industry and explain that strategic decisions relate to the selection of energy production technologies, network configuration, supply and demand contracts, and sustainability issues, while tactical decisions are more related to production scheduling and inventory decisions. A major component in the design of these networks is uncertainty, and therefore, stochastic programming is a modeling tool of choice in many of the literature work.

2.1.2 Crop Selection in Agricultural Production Planning

The problem of optimal land use under the context of agricultural crop planning has also received limited coverage. The earliest works found regarding the crop selection problem for a geographical region was in Boer and Chandra (1978) who use a series of Cobb-Douglas production functions to determine the influence that varying levels of required labor and effort for producing particular subsistence crops has on a region's actual crop selection, labor usage and cash income. They then use these functions to explain production patterns in Fijian and Indian farms. In a later work, Bocco et al. (n.d.) indirectly address the problem of determining alternative production options by using simple goal programming and a metaheuristic with a land and capital constraint applied to horticultural companies in Cordoba, Argentina. However, the models used in this work are simplistic and lack detail in important components, such as stochasticity, labor and environmental constraints, and planning.

More recent works found in relation to the crop selection problem also use goal programming and metaheuristic techniques to determine the best set of crops to plant in a particular land. Examples of this type of work include: (1) Brunelli and Lücken (2009) who propose multi objective evolutionary algorithms to select among five crop alternatives the ones that best fit the soil characteristics as a way to minimize the need for soil treatment, reduce cost, and reduce potential environmental damage. (2) Pal et al. (2009) propose a hybrid algorithm that combines fuzzy goal programming, interval valued goal programming and genetic algorithms to allocate crops to a particular land depending on the available supply of resources. (3) Chetty and Adewumi (2014) apply swarm-based metaheuristic algorithms to make recommendations on crop planning based on existing irrigation schemes.

2.2 Integrated Assessment Models

Another area of interest pertaining to the problem in this dissertation pertains to the assessment of the production capabilities of particular regions. As noted earlier, the need to close the agricultural production yield gap in order to satisfy the future demand gap has motivated a lot of research in the area of yield gap estimation, land exploration, and the assessment of future agricultural production patterns. Many of these assessments are done at the macro level and are geared towards determining productivity potential. Agricultural development and land use is undertaken to satisfy a number of socio-economic goals from different interest groups, such as production, employment, and profit, as well as environmental stability, pollution, abatement and political compensation (de Wit et al., 1988) and therefore, many of the studies in this area is geared towards simultaneously satisfying different objectives. In this research area, multiple goal programming has become a standard for research tools aiming to decide on feasible development pathways within a wide range of technical and socio-economic scenarios.

Within the research in integrated assessment models, there are several examples of this type of models, including IMAGE (Image Model to Assess the Global Environment), GCAM (Global Change Assessment Model), MAgPIE (Model of Agricultural Production and Its Impact on the Environment) and REMIND (Dinar and Mendelsohn, 2011). These models are very similar to Bio-Economic Farm Models that link formulations describing farmers' resource management decisions to formulations that describe current and alternative production possibilities in terms of required inputs to achieve certain outputs and associated externalities (Janssen and van Ittersum, 2007). One should also note that these models are geared towards crops, such as wheat, beans and other types of grains that do not have immediate perishable characteristics as compared to fresh produce. Also, the simulation and linear programs developed in these integrated assessment models, such as in IMAGE and MAgPIE, analyze crop production at a grand scale over long periods of time. Most of these assessments do not consider technologies that are deemed more sophisticated, more realistic and have the ability to push yields above the levels reported by yield potential assessments, which are often assumed non-manageable in analysis, such as greenhouses and high tunnel structures (Dietrich et al., 2014). Therefore, these tools lack the level of resolution needed to consider different labor, resource, and logistic variables required to make accurate estimations of profitability in targeted technological investments within a geographical region of interest.

2.2.1 Land Use Exploration and Management

Land use and exploration models have been a topic of interest for the past few decades. Land use models often apply recursive dynamics with myopic agents and base trade decisions either on historical patterns or on cost minimization (Dinar and Mendelsohn, 2011). A common approach in many of the academic works in this research is the integration of goal programming and linear programming and simulation. One of the earlier works was performed by de Wit et al. (1988), which uses interactive multiple goal programming to address different competing goals in agricultural development. The interactive approach requires that the desired solution is attained at the end of a series of iteration cycles. The first cycle begins with a feasible space in which the minimum value of each objective is satisfied, and then, each goal is maximized independently, while setting the rest of the goals as minimum requirements. In these iterative goal programs, the resources in the region include the area and quality of various land types available, available labor, additional labor hired outside the region, endowment of capital goods, etc. The quantity of these resources are externally simulated and used as inputs into goal program at each iteration. However, it does run into the difficulties of goal programming, which is determining the weights attached to each goal, as well as a limitation on post-optimal analysis (Glen, 1987), and decisions are myopic.

In one of the earliest works found related to this topic, Barnett et al. (1982) use goal programming with multidimensional scaling to model a Senegalese subsistence farmer's decision problem. The authors point out that a multi objective approach is particularly relevant for subsistence farmers because they are frequently said to possess conflicting objectives such as profit maximization, risk avoidance, and maintenance of minimum food requirements. Another early work and one of the most complete works taking this approach was performed by Bouman et al. (1998), who determine sustainable land use options in the North Atlantic Costa Rican region. In this work, the authors quantify the trade-offs between socio-economic and biophysical sustainability at the regional level. The land evaluation was used to determine which crops and pasture maximize the profitability of agricultural operations in the North Atlantic Costa Rican region. In this study, the parameters of crop and livestock conditions are simulated and then used as inputs into a linear program. The objective function of this program is the maximization of regional economic surplus through the selection different land use alternatives subject to land, market, and labor resource constraints. For the set of market constraints, the linear program considers base price and quantity for the products considered within different regional and international markets. Additionally, the study considers 72 alternative technologies with two levels for pesticides and mechanizations, in which for each combination of technologies there is a target production levels ranging from certain minimum levels to maximum attainable production. Although, the study itself is complete and allows for the assessment of different policy-based options, it does look at the problem from a macro level perspective, again, assuming that farmers will trade according to given simulated model parameters. Furthermore, the linear program itself is static and only considers a single period.

In similar fashion, Zander and Kächele (1999) use multiple goal linear programming (solved through a Simplex method) and trade-off functions to make land use decisions given different agricultural, economic and environmental objectives. The operational and tactical decisions with respect to growth of a specific crop at a specific site are simulated in order to integrate competing goals. Similarly, Wallace and Moss (2002) employ recursive goal programming to track the adjustment process of six dairy and beef/sheep farms. In this case, the authors iteratively solve a goal program in a sequential manner as price expectations and resource vectors are updated. Dogliotti et al. (2005) use process-based simulation models with empirical data to quantify inputs and outputs of production activities. They then develop a multiple goal linear program to allocate production activities to a farm with land units differing in soil quality, while maximizing socio-economic environmental objectives. Seven objective functions are optimized one at a time; when one of the models is optimized, the others are used as constraints. The ultimate purpose of this work was to design improved farm systems for seven farmers in South Uruguay. Additional examples of simulation and goal programming based solutions to assess agricultural development of particular global regions Lu et al. (2004), Gibbons and Ramsden (2008), and Dietrich et al. (2012).

2.2.2 Estimating and Closing the Yield Gap

Another closely related topic to this research is the estimation of the true agricultural yield gap that exists within a particular region. This topic has gained attention over the last decade, as global consumption forecasts hint that current agricultural productivity is simply not enough to cover this growth, and therefore, there is wide interest in estimating the magnitude of the productivity increase needed to cover this gap. Van Ittersum et al. (2013) argue that determining the yield gap is critical for four main reasons: (1) it allows policy makers to determine the most important crops, (2) it enables effective prioritization of research, development, and intervention, (3) its results become a key input to other economic growth models, and finally, (4) it aids in the evaluation of the impact of climate change and other future scenario that influence land and natural resource use. However, the development of accurate and consistent yield gap estimations within different regions has proven to be a hard task due to a limitation on the availability of data at both the regional and global scale (Neumann et al., 2010).

One of the main difficulties of yield gap estimations is that assessments are mostly dependent on the specific conditions of a particular region. For example, in semi-arid cropping regions, it is particularly important for farmers to maximize grain yields in seasons when the rainfall is adequate to produce profitable crops (Anderson, 2010), and thus, its yield gap estimation would need to consider this fact. Similarly, differences in grain production efficiencies may be correlated to irrigation, influence, market accessibility, agricultural labor and slope (Neumann et al., 2010), and thus, if one considers various regions at a time, a yield gap estimation methodology tailored to one region might give different results in another. In current literature, there are at least four methods for estimating yield gaps at the local level: field experiments, yield contests, maximum farmer yields based on surveys, and crop model simulation (van Ittersum et al., 2013). However, much work has yet to be done in this area, as it is a relatively new research direction.

The latest research in this area has been geared towards the development of methodology frameworks. For example, Dietrich et al. (2012) develop a land-use intensity measurement, which attempts to combine R&D, infrastructure, and managerial components to assess current agricultural productivity. In another example, Hoang (2013) present an analytical framework to analyze the production efficiency of different agricultural systems and illustrate their work through an empirical analysis of rice production in Sri Lanka. Neumann et al. (2010) use econometric-based models along with spatial analysis to explore maximum attainable yield, yield gap and efficiencies of wheat, maize, and rice production. Tao et al. (2009) construct a model through which they can examine the impacts of climate variability on crop phenology and yield over a large area. However, estimation of yield gap is only one component of the whole problem. The other component is the development of strategies aimed at closing the yield gap, which still has ample room for improvement.

Verburg et al. (2000) outline three primary sources of improvement on regional production yield: increases in agricultural inputs such as irrigation and fertilizers, increases in production efficiency and technological change. Also, tactical management, including the choice of crop and cultivar, fertilizers, and weed, insect and disease control can have great impact on productivity, especially when combined with improved management of strategic factors (Anderson, 2010). Similarly, in more advance agricultural systems, understanding when a system has reached its yield plateau (i.e. point at which the marginal profits from investing in new technologies is relatively small) is critical to determining whether it is possible to resume yield advance, or if the focus should be placed on accelerating yields in other grain producing regions (van Wart et al., 2013).

2.2.3 Projecting Future Land Use Patterns

As part of the integrated assessment models, a key component is the ability to accurately forecast future land use patterns. Again, for most of these models, the general approach is to use a combination of simulation and basic linear programs. The major inputs into these models are simulated parameters for random variables such as climatic conditions and production yields. Then these are fed into an LP program using goal programming techniques in order to incorporate multiple objectives, such as regional profits, environmental goals, social impact, etc. Furthermore, the majority of these models incorporate recursive properties in order to project the simulated events into the future. A prime example of these kind of models is MAgPIE first presented in Lotze-Campen et al. (2008), which is a recursive dynamic LP that attempts to integrate economic and environmental processes. MAgPIE works on a time step of 10 years in a recursive dynamic model, while the link between two consecutive periods is established through land-use patterns. The crop yields for each region is supplied externally by a simulation model (called LPJmL) which simulates the vegetation process, including climate and soil conditions, water availability, and plant growth, as well as CO_2, temperature, and radiation on yield directly into account. Trade in food products between regions is simulated endogenously and is constrained by minimum self-sufficiency ratios for each region. The integration of MAgPIE and LPJmL is the basis for other types of analyses, such as analyzing future trade scenarios in Schmitz et al. (2012), adaptation to climate change through choice of cropping systems and sowing dates in Waha et al. (2013), forecasting technological change in Dietrich et al. (2014) and projecting future crop productivity in Müller and Robertson (2014).

2.3 Resource/Technology Allocation and Profit Sharing in ASCs

The second phase of this research effort studies the efficient allocation of technologies and resources to farmers in marginalized regions, while considering the mechanisms through which profits could be shared. Access to this kind of decision-making tools would provide small farmers with the means to access price information, receive adequate technologies, and connect directly to market opportunities, which often limits their true commercialization capability (Markelova et al., 2009). In this review, we highlight the most relevant work in both of these research areas.

2.3.1 Allocation of Resources

In one of the earlier works pertaining to the adoption of technologies in the agricultural setting, Just and Zilberman (1983) develop economic-based models that primarily explain different land-use allocation and technology adoption structures within an agrarian set-up . The authors used a mixed-integer program to represent the decisions made by a particular farmer that decides whether to allocate his land to a traditional technology or to incur a fixed cost for a new technology (in which case he can allocate his land in any proportion between the two technologies). Similarly, Carter (1987) use non-cooperative, Cournot-Nash equilibriums to examine the emergence of parcellation in some agricultural industries. As part of the study, authors examine different alternatives to the sharing of resources and distribution of income among farmers in agricultural regions. In a later work, Zijun Wang et al. (2002) explain and examine the moral hazard problem that sometimes obstructs equity financing of farm businesses. In this case, authors use a structure in which a supplier of external equity capital cannot directly observe the partnered farmer's effort but can observe the random outcome of the effort. The authors then use the model to solve for the optimal farm income-sharing rule that maximizes the effort provided by farmers, which in turn increases overall net returns.

In one of the most complete works dealing with small holder farmers' access to market, Markelova et al. (2009) draws on case studies to understand how collective action can help address inefficiencies, coordination problems or barriers to access for smallholder farmer market access. Authors argue that most cases of successful collective marketing highlight the crucial role of a facilitator who catalyzes collective action, provides information and technical assistance, and builds capacity of a group to effectively engage in marketing activities. While there is literature consensus on the importance of facilitators, not so much on who is best positioned to take on this role. If the approach of 'linking farmers to markets' is to be successful, its proponents need to accept commercial realities and not prioritize the poverty reduction goal at the expense of business sustainability (Markelova et al., 2009).

2.3.2 Vertical Coordination and Profit Sharing in ASCs

Within vertical coordination and revenue sharing literature, there are some works that relate to agricultural supply chains. An application of these works is observed in the area of renewable resources, which has similar random components as one would observe in an agricultural setting (e.g. yield, price, etc.). For example, Bai et al. (2012) propose a game-theoretic model that incorporates farmers' decision on land use and market choice into the biofuel SC design problem. In another example, Yue and You (2014) develop a deterministic MINLP to simultaneously optimize operational decisions and profit allocation mechanisms in supply chain optimization, namely transfer prices and revenue share policies (by using a Nash type function) among supply chain participants of cellulosic bioethanol materials. The authors propose a solution strategy based on logarithmic transformation and branch and refine algorithm for efficient global optimization of the resulting non convex MINLP problem. In this case, profit shares received by each actor in the supply chain are formulated as a constraint that makes sure that profits are shared equally among the players in the supply chain.

There is also more literature work related to revenue sharing within more general supply chains for products that are not necessarily perishable. For example, Gjerdrum et al. (2001) present programming formulations for fair and optimized profit distribution between members of multi enterprise supply chains, arguing that simple maximization of supply chain profits does not automatically and fairly apportion it fairly among players. In Nagarajan and Sošić (2008), authors give a review of works related to cooperation among supply chain agents with an emphasis placed on profit allocation and stability, including description of the set of feasible outcomes of cooperative bargaining models commonly observed in supply chain management. Authors show the application of the bargaining model to profit allocation/sharing among a group of two or more agents and explain that this model can be solved using Nash's Bargaining model, while giving a unique solution satisfying Nash's equilibrium axioms, ensuring it symmetry (i.e. no player is better off deviating from the solution) and feasibility.

2.3.3 Contract Farming

Within the area of revenue sharing mechanisms and contract allocation, an emerging and interesting practice within developing countries is contract farming, and although it might not be a new approach, it is an innovative undertaking for most smallholders and for many buyers in developing countries (Will, 2013). In this set-up, corporations engage in forward contracts with farmers, while supplying the needed inputs in exchange for a quality controlled product available in a time manner for the purpose of ensuring quantity and price stability (Huh and Lall, 2013). These types of schemes are usually organized by large-scale processors, exporters, or supermarket chains (Minot, 2007). Key and Runsten (1999) note that in addition to raising income of growers, contract farming may also create positive multiplier effects for employment, infrastructure, and market developments in the local economy; however, members of rural populations *do* run the risk of limited gains or even exploitation. In the more modern applications of contract farming have been under food processing companies contracting with a large number of relatively small and financially challenged farmers (Federgruen et al., 2014).

Huh and Lall (2013) take a look at the decision problem of a farmer associated with allocating his land among different crops with varying crops with varying water requirements, under the assumption that a subset of crops may be associated with a forward contract offered by a buyer before a season begins. The problem includes a decision to acquire a certain amount of irrigation water capacity prior to the season and allocating this capacity as needed during the progression of the season based on the crops selected. In a similar approach, Federgruen et al. (2014) take the perspective a single manufacturer with a collection of manufacturing plants in the production of potato chips. In this case, the manufacturer offers a "selected" group of potato farmers a menu of contracts (i.e. quantity-price combinations) before the growing season, and then afterwards, attempts to minimize the distribution costs from the selected farmers to the production facilities. In their model, the only random factors affecting a farmer's total yield is water supply; during the season, rainfall and the well capacity of the farmers are revealed. Furthermore, the authors formulate the problem as a Stackelberg game, where the manufacturer is the leader and farmers are the followers, where the assumption is that joint demand and rainfall distributions are known.

Similarly, Tan and Çömden (2012) present a firm whose objective is to match random supply of fruits and vegetables that survive only one growing season from a number of contracted farms. The main optimization problem is to determine the farm area and the seeding time of different farms producing a single product over a planning period. The formulation uses the length of the maturation and harvesting period as random variables to determine the probability that the output from a particular farm will be available at a particular period, as well as its total supply. Consequently, the authors construct the situation as newsvendor problem with random yield. In this case, the assumption is that firms make the necessary investments to complete harvest without any resource conflicts and that resource management can be effectively managed once long-term decisions have been made. Similar to our current problem, the application considers a company that supplies agricultural products to retailers throughout the year and thus must use farms located in different regions based on the time at which they can start production.

2.4 Solving Schemes within Stochastic Decomposition

The final section in this literature review pertains to the solving mechanisms addressed in this work. Specifically, the integration of machine learning techniques within a stochastic decomposition. From a review of literature, the amount literature dedicated to the use of machine learning models within a stochastic optimization framework is limited. There have been works geared towards the organization and reduction of second-stage scenarios that aim to organize them such that they can be sampled within the stochastic framework. Other types of works aim to use statistical tools, such as response surface methods, to seek scenarios that provide the most information from which to sample. Finally, there are also works that aim to learn from the intermediate outputs to make first-stage recommendations. With regards to scenario construction, there has been some work done in the area of organizing scenarios to solve the stochastic optimization problem. For example, Chen and Mehrotra (2007) use a quadrature algorithm to create a sparse grid from which to sample scenarios within a multi-stage framework. Bailey et al. (1999) apply response surface techniques to approximate the objective function of the two-stage problem with recourse through which provides insights as to the direction of the maximum and minimum sensitivity to changes in first-stage variables. Bayraksan and Morton (2011) develop a sequential sampling procedure in which the number of samples is augmented or resampled according to the assessment of current solutions. Similarly, Chung and Spall (2015) incorporate techniques from statistical experiments design, specifically a response surface methodology. In this case, the authors divide the stochastic problem into two parts, which includes an exploratory phase, during which a fraction of the time is reserved for conducting informative measurement, assuming a level of control over the amount of information collected within the probabilistic space.

The use of machine learning techniques to learn from a stochastic problem's results is another interesting aspect of this dissertation. Defourny et al. (2012) propose a hybrid strategy based on machine learning techniques for generalizing to nonlinear decision rules. In this case, the authors solve an optimization on a scenario tree to obtain optimal decisions and use these decisions at all stages in the tree to fit the policy function from which the authors infer decision policies from the solutions of the scenario tree approximations to multi-stage stochastic programs. Similarly, Wang et al. (2008) integrate a genetic algorithm to estimate the expected profits form of all scenarios within a stochastic sampling procedure and apply the methodology to optimize capacity planning and resource allocation under variance of different demands and expected returns in long-planning horizon.

2.5 Conclusion and Literature Contribution

In this review of literature, we have focused our literature review on three basic lines of research most relevant to the problem of this study: agricultural production planning, land use exploration and integrated assessment models, and profit allocation/incentive mechanisms within agricultural supply chains. Overall, we have observed that all three lines of research have received an increasingly amount of coverage over the last few decades, as the agricultural industry has become an ever more focal point in today's society. However, despite the large amount of interest received, there are still literature gaps that should be addressed.

Current literature in agricultural production planning is quite extensive. As previously discussed in this review, there is large amount of research regarding tactical and operational level planning, while literature regarding strategic level planning is a bit more limited. Also, available models addressing the crop planning problem are mostly dependent on metaheuristics, which often lack interpretability of solutions, which renders them difficult to use in applied settings. Specifically, little to no research has been devoted to the strategic sourcing and planning of agricultural production based on the environmental and resource characteristics of different geographical regions, supply chain decisions, and market conditions.

Land use exploration and integrated assessments is another topic that has garnered attention in agricultural related literature. In this case, general assessments have been performed on the difference between actual and potential production. Also, the majority of research has focused on regional assessment models for purposes of policy making and future projections of land use. However, these assessments take a generic approach into how future land use will take shape and cannot be really be used by individual farmers to make strategic and tactical decisions into where to source products and when to produce them given temporal environmental traits and technological investments made by the main farmer.

The integration between statistical analysis and machine learning tools to capture information derived from a stochastic decomposition problem is a topic that has not received much coverage in literature. The use of response surface methodologies, as well as learning tools, have been applied to structuring the sampling procedure added to the optimization model. In this dissertation, a stochastic decomposition design is provided that focuses on the overall first-stage solution and assesses the stability of first-stage solutions. In this case, the objective is to learn the relationship between first-stage solutions and the scenarios sampled during the second-stage and to use this information to gain insight to the stability of specific first-stage decisions.

The development of technological/resource allocation and profit sharing mechanisms among vertically coordinated echelons of agricultural supply chains is a topic that has received increased coverage over the last few decades. The research to be presented in this dissertation will contribute models that seek to allocate and position needed technologies and resources within identified regions such that profitability requirements of both investor and individual farmers are met. Within this research context, the research will also investigate the applicability of using different profit sharing mechanisms, such as auctions, to elicit pricing information from the different farmers in order to coordinate production, such as those proposed in Mason and Villalobos (2014).

The modeling tools resulting from this dissertation seek to increase the commercialization capabilities of farmers in marginalized regions. By providing established farmers the ability to identify regions with hidden production capabilities, they can expand current production windows through which they can service the market and in turn allows them to tap into some of the profit margins are currently claimed by supply chain intermediaries of the agricultural industry (Section 1.3). For marginalized farmers, this means a direct access to global agricultural markets, as well as the benefits associated with higher income to regional economies. Lastly, these tools can be used to supplement current policy assessment models by increasing the resolution of the farmers' decision-making process searching for alternate sources of production.

3. METHODOLOGY

3.1 Background into Traditional Agricultural Planning

The ultimate purpose of the research presented in this dissertation is to develop decision support tools through which farming organizations can identify geographical regions with hidden potential to produce high-value products and integrate them into established agricultural supply chain (ASC) in a sustainable manner (i.e. can be implemented without long-term government or social intervention). As previously stated, we hypothesize that within many of these marginalized regions, there exists alternative production options that could not only be more profitable for farmers in the region but can also attract the level of external investments needed to upgrade their production capabilities.

There are several aspects that need to be considered when assessing the profitability of investing and integrating a particular region into an ASC. Since a region's agricultural production is often very dependent on its environmental and labor characteristics, as well as its historical demand and access to market, farmers in these regions seldom deviate from traditional agricultural practices and logistics restrictions. In this case, the set of products commercialized by these farmers rarely change given limitations set by the region's characteristics, such as temperature, rainfall, etc. Consequently, the set of markets that these farmers can target is also limited to the set of products they can produce. In many cases, changes in strategic planning are long-term and tend to be reactive to changing circumstances rather than a proactive in order to avoid losses.

Interestingly, tactical level planning tends to be much more dynamic due to the variable nature of the industry. Traditionally, the farmer constructs an agricultural plan based on the beliefs of a region's upcoming seasonal weather patterns, as well as his expectation of future market prices for a set of traditionally cultivated products. Most farmers also set up a set of contracts for the amount of products and the price at which supply is needed. Based on this assessment, a seasonal tactical plan is developed in which production quantities are determined in order to meet supply obligations, based on information on expected available resources, market signals, and environmental conditions (Figure 3-1). The major constraints to production derive from the amount of available resources, as well as the environmental characteristics of the region. Thus, for most traditional farmers, production is strictly seasonal.

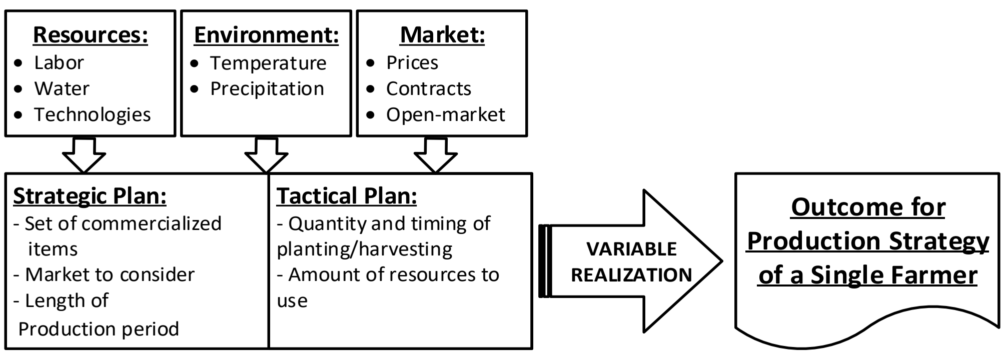


Figure 3-1: General Inputs into Agricultural Production Planning

Previously, we argued that certain regions have the environmental conditions that would allow production of high-value products, if adequate technological upgrades are made that could extend production windows and/or protect against regional climate variability. In some instances, the profitability of a market opportunity for horticultural products may be high enough to garner investments in the technological upgrades of a particular region. For example, an area might have a production window in terms of temperature that is slightly too short or too variable to produce tomatoes and in turn settle for a lower-value crop, such as wheat or sorghum. Without access to proper technologies, marginalized farmers will most likely also observe large yield gaps. In this example, since sorghum is a crop that generates low profit margins, it may also be difficult to incentivize external investment for the upgrade of their production system and help them close the yield gap. In this case, the ability to identify those marginalized regions that could potentially produce high value crops may be a catalyst for more sophisticated agricultural planning tools.

3.2 Envisioned Framework of Study

In this case, we seek to provide the mid-to-large holder farmer with better strategic and tactical level planning tools that would allow him to better identify regions with production potential, as well as the means of incorporating these regions into an ASC. In this section, we explain the basic framework of the basic interactions between the main farmer and those in marginalized regions considered in the research presented.

3.2.1 Possible Scenarios to consider

We begin with the traditional two-way interaction between the farmer and a particular target market. Let us consider the case in which we have an agricultural zone and a specific market at a given convenient time period $t \in T$ (Figure 3-2). For instance, the weeks of the year the farmer can efficiently reach a particular market. Traditionally, the supplier plans his planting/harvesting according to the time-dependent environmental and resource characteristics of the region (e.g. temperature, sunlight, precipitation) which are constrained by fixed capacity and resource limitations, such as water, labor, and infrastructure. Overall, the farmer in this region seeks to maximize overall profits, while reducing production costs and meeting contracted (or open-market) demand.

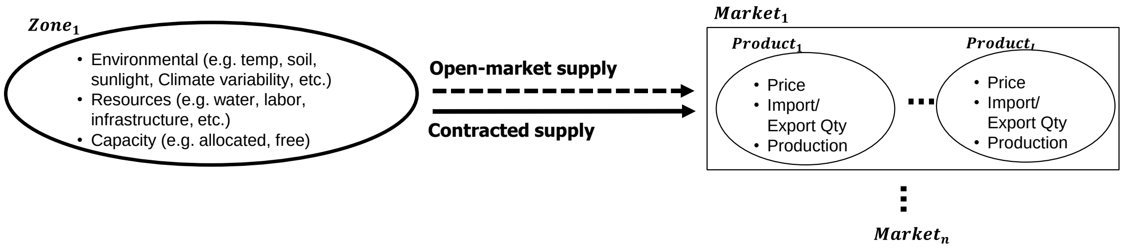


Figure 3-2: Basic Buyer and Supplier Interaction at Time t

Most traditional production systems are represented by this simple two-way case. However, there are some difficulties and limitations that arise from this simplistic set-up. For instance, the farmer must match the production in his region constrained by time-dependent environmental and resource variables to the time-dependent demand patterns of the market place. For example, producers in Yuma are limited by the high temperatures of the summer months, while producers in Michigan are constrained by the frigid winter conditions. Therefore, the maximum amount of production that these farmers can obtain is limited by the amount of production he can produce during the season and the environmental conditions of his region. This limits the capability of the farmer to capture new market opportunities that will most likely arise in the future as consumption for specific agricultural products increase. Secondly, since agricultural production is mostly regional and weather dependent, when a damaging climatic event hits a particular region, it has the capability of destroying entire harvests resulting in major financial losses for the supplier.

Now, let us consider the case in which a farmer (e.g. established local food system) identifies and incorporates zones with production potential into his supply chain for the purpose of diversifying operations (Figure 3-3). These zones may not be within the same geographical area as the main farmer and therefore could have their own set of environmental and land characteristics that can complement the main farmer's original region. However, depending on the proximity of these regions, they may also have some overlapping environmental, resource, and land characteristics with the main farmer. Furthermore, these regions may lack the funding, technologies or access to the market that would allow them to market their product or to efficiently plant/harvest high-value products. In this case, it might be in the best interest of the main farmer to invest in the technological upgrade of these regions and incorporate some of its farmers them into the ASC. Among the underlying questions in this scenario is how to identify these potential zones and how to incorporate them into the main farmer's operations. This includes optimal designs of the supply chain such that transportations costs are considered within the profitability assessments.

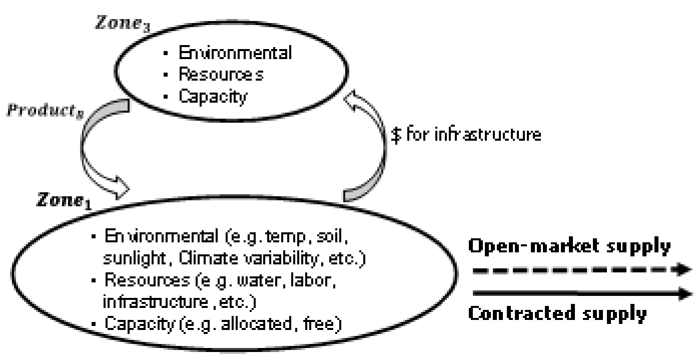


Figure 3-3: Incorporation of Other Zones

Before we continue, it is important to note some of the flexibilities that such operation provides. For example, let us assume that the main farmer observes an increasing trend in the consumption of a particular Product B within a market to which he already has access to but his current production capability (due to seasonal limitations, resource constraints, etc.) prevents him from producing. One of the options for this farmer is to identify a region with the necessary environmental and resource characteristics to produce this item. However, this region might also need additional investment in infrastructure and technologies to upgrade their current production system to produce Product B. In this case, the main farmer and the farmers in this identified region may find it mutually beneficial to reach an agreement in which the former is able to expand his product basket by investing in this other region, while the latter is able to not only expand her own production capability but may also receive higher profit margins.

These zones also provide possible options for better risk management, and we present this by the use of another instance of this problem. Let us assume that the main farmer's production of Product A is being increasingly hampered by a higher frequency of extreme events. Now, this particular farmer may benefit from identifying and incorporating those zones that also have almost the right environmental and land characteristics to produce Product A, but may lack the production window duration for producers in the area to plant/harvest this product, as well as the access to market that the main farmer would have. In this case, these zones might be able to produce Product A, if they had additional infrastructure that would help extend their production window. In this case, the main farmer may find it beneficial to invest in the identified region's infrastructure in order to broaden and pool some of the risk associated with volatility of climatic events in his own region. The farmers in the identified region might also be benefited from better infrastructure and technologies, as well as a buyer for the products they produce.

One should note that for the cases above, the difference in negotiation power between the main farmer and those *other* farmers in these identified regions would most likely be large.

This is important since it would be naïve to assume that once the farmers in this region receive this new technology and/or infrastructure, they will not simply end ties with the main farmer and enter the markets by themselves. The farmers in these identified regions would most likely be farmers that are marginalized or may only be producing low profit margin products and that may not have access to the market or to a specific portion of it. Furthermore, as we move to the next phase of the problem, the level of transparency in the sharing of information, as well as the efficient allocation of resources, would have to be considered during the actual implementation.

3.2.2 Allocation of Resources and Technologies

The result of the first phase of the problem is a group of identified regions that under targeted technological investments could produce higher-value, horticultural products. The question in this case is how to best allocate technologies to farmers within identified regions, in such a way that all parties are satisfied. For the main farmer, the overall objective is to minimize the costs of taking the product from the identified region into the market place. On the other hand, the individual farmer must weigh her own incentives of transitioning over to the production of new products and determine the level of return to be expected in order to make this change. Also, important within this context is how to assign the usage of these technologies to the horticultural products to be produced. For example, there may be a case in which a particular product A has been determined to have higher profitability in the market place over product B over the same period of time. In this case, it would be desirable for the technological infrastructure to be allotted for the production of product A to be produced by a certain set of farmers. This might also include providing fertilizing material in order to modify soil chemical characteristics and change the quantity of nutrients and acidity in the ground making agriculture possible (Johnson et al., 1991).

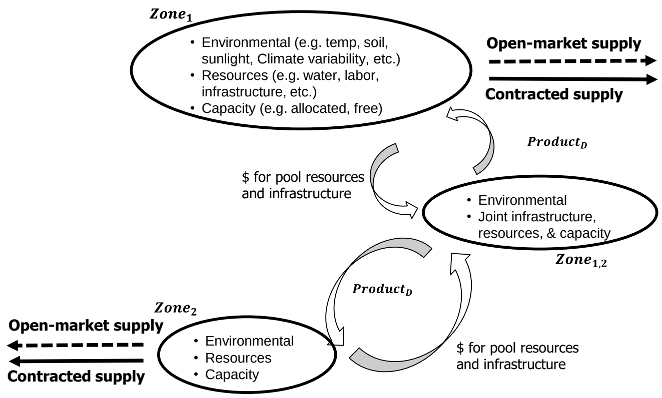


Figure 3-4: Pooled Infrastructure and Resources

Another aspect of this problem in the case in which the investment capital required is either too high or too risky for any one farmer to take. In this case, it might be in the best interest of both farmers to enter a joint venture by sharing the upfront costs of an infrastructure investment or in the development of the production capabilities of a particular region. However, this adds another level of complexity to this problem, since investment costs are shared and operations and resources would most likely also be shared. In this case, this aspect of the research is outside of the scope of this dissertation is left as future work.

3.2.3 Overall Hierarchical Plan

The set of decisions regarding the actual allocation of technologies within an identified region would most likely influence the initial set of region identification decisions. Thus, it is important to construct a modeling framework through which the decisions on which regions to incorporate can be improved upon by the set of decisions on how to allocate technologies within the region. In Figure 3-5, we describe this point in more detail. Market assessment models are based on information of product price and demand patterns, while the region selection model is based on information of the available resources, environmental characteristics, and accessibility. With the use of this data, decisions on which crops, markets and regions to integrate to the ASC can be made. This set of decisions includes the type of technologies to be implemented, the design of the supply chain, and any additional resources that would need to be incorporated.

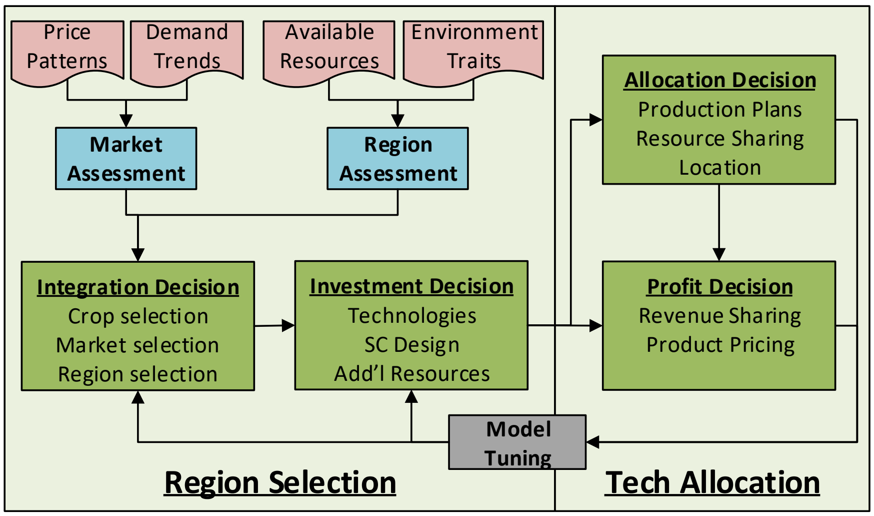


Figure 3-5: Overall Hierarchical Plans

Once a particular region has been identified and the investment requirements are determined, they become inputs to the next phase of the problem, which is the set of decisions regarding the allocation of technologies to farmers within a particular region. This would be planning of production that use this technology, which includes what type of horticultural crops the produce and how should these new resources be shared. Also as part of this process, it is important to consider different mechanisms through which revenues will be shared once technologies and resources have been assigned. This includes auction mechanisms such as those proposed in Mason and Villalobos (2014). The proposed technology allocation formulation would allow the implementation of more sophisticated revenue sharing mechanisms. Finally, the results from this model can be used to refine parameters used in the region selection phase.

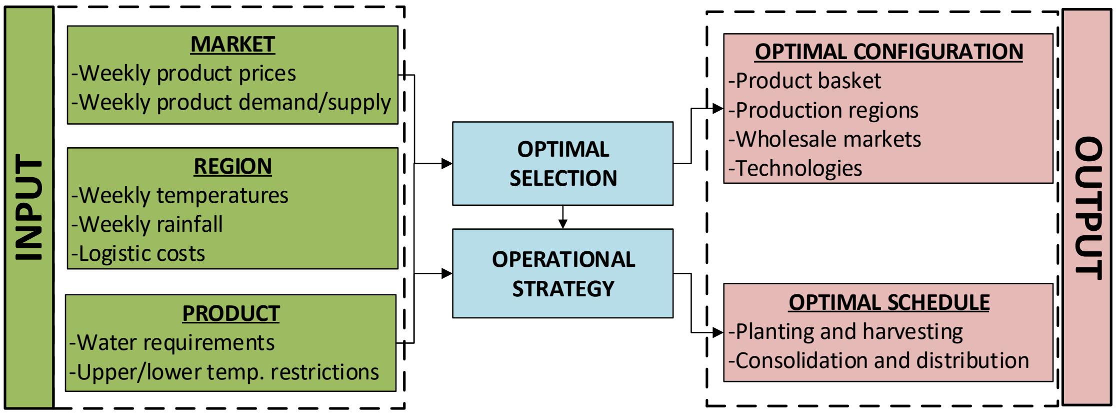


Figure 3-6: Hierarchical Planning

Access to data will be an important aspect of the research as the information needed is based on both market, as well as regional geographical information. Figure 3-6 presents the type of information that would be included in the region selection and technology allocation models. The accuracy of this information can have an impact on the results of our models, since stochastic programs are dependent on the expectation values of random variables.

3.3 Objectives of the Research

In this dissertation, we aim to develop a modeling framework that attempts to connect demand opportunity with production potential within under-producing regions. In order to achieve this, we target three main objectives:

1. Development of a strategic planning model that captures an initial deterministic version of the parameters and variables. The goal of this initial model is to be able to select the regions that could potentially produce items that show potential

profitability in a particular market given temporal environmental variables. The set of variables include a region's temperature across time and crop production yields, as well as the environmental and resource requirements of each particular crop. On the other hand, this initial model also considers market-based parameters, such as price and consumption for a particular set of items across time. This includes those regions that have the hidden potential to produce items, only with the introduction of certain technological upgrades.

2. Expansion of the initial deterministic model to consider the variability of stochastic parameters. The aim of this stochastic model is to attempt to increase the reliability of the solutions by considering variability that is not considered in a deterministic model. As part of this phase, the effects of different types of yield functions on the feasibility space can be explored.
3. Development of a model that seeks to allocate technologies and resources to farmers *within* an identified region, in which the farmers will have to decide based on their own level of profitability whether to transition to a new set of higher value products. Similarly, the entity making the investment (referred to as to the central, decision-maker) would have to decide the physical assignation technological infrastructure resource and the product that each farmer will produce based on the production targets set by first phase results.

In this research effort, a modeling formulation is proposed that addresses each of the objectives described above. In order to validate our proposed models, we are planning to collect data from farmers in the region of the Mexican state of Sinaloa, as well as Yuma, AZ, and use this information to a construct more case studies and solutions.

3.4 Phase I: Deterministic Model for Region Identification

The first phase of the study consists of determining alternative production options for a group of geographical regions. The alternative options will be given by a set of selected horticultural items estimated to have market potential. In this case, we propose the development of optimization models that consider environmental (e.g. temperatures), resource (e.g. water, labor), and technological parameters (e.g. greenhouse, fertilizers, etc.) of different geographical regions, as well as their accessibility and supply chain costs, to assess "hidden" production capabilities. The model attempts to match these characteristics to the resource and environmental requirements of particular crops. The initial overall objective will be to maximize the expected profitability from investing in the technological upgrade of an optimal selection of geographical regions and alternative production options. As part of the proposed modeling tools, we will also consider the randomness of some of the parameters, such as temperature and rain precipitation, associated with each region during different time periods. To solve this problem, we propose the development of two-stage, stochastic programs to address this problem. We will also explore ways through which we can exploit special modeling structures to find efficient solution methods. The overall objective is the maximization of profits over all regions and crops: the first summation set details the revenues received from product sold at market and salvaged, the second set are the costs associated with the investments on technologies, the third summation sets are the costs associated with planting and harvesting, the fourth and fifth summation sets are the costs associated with water and labor allocations, while the last set of summations are costs associated with the movement of the product from the harvesting field to the market place.

One of the key contributions of this phase of the dissertation will be the consideration of yields as a function of different technology decision variables and environmental conditions. As part of this research, we will investigate ways through which yields can be estimated under different conditions. This will allow us to determine the optimal configuration of technology and geographical locations that maximizes profits. Each of these different configurations alters the production conditions of the crops, which in turn may modify current production windows and provide the opportunity to produce alternate crops. For example, as shown in Figure 3-7, the implementation of high tech irrigation systems combined with greenhouse structures in Region B may give higher yields and extend the production window for a particular crop. However, based on market behavior, this might not necessarily be the best choice, as it might just be as profitable to have normal irrigation technology on open field in Region A for the same crop. In the stochastic case, the investments on improved technologies may actually reduce yield variability and further alter the optimal selection of location and investment technologies, as it will be discussed in the next section.

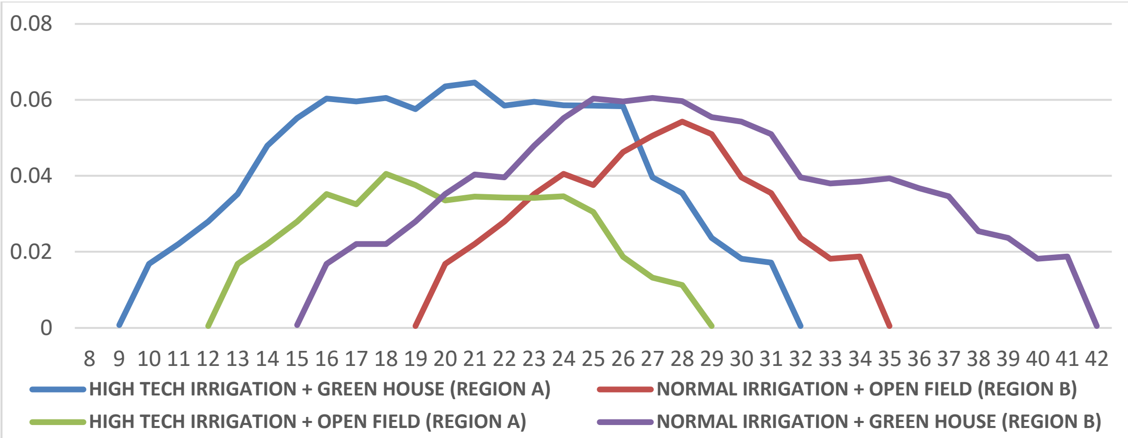


Figure 3-7: Expected Yields under Varying Location and Technologies

To obtain yield estimations of crops under varying technologies, we will interview farmers and agronomics experts in agricultural regions with current production of high-value crops and use this information to infer values of alternative production windows. However, as we progress in the research, we will also explore alternative representations of the yield as a phenological functions of different growth variables, such as temperature, sunlight, water, fertilizers, etc. In this case, yield estimations will be altered directly by the selection of technologies and locations.

An important consideration of the research is the cost estimation of implementing technologies within a particular region. Since investments are large and the period of evaluation have long-term horizons (e.g. 5-year, 10-year), we would need to adjust the amortized cost of the total investment to the planning period used in our model. Also important are the estimation of logistic costs and supply chain designs that may alter the solution space. For example, the main farmer may find a region non-profitable based on high logistic costs associated with the current location of distribution centers in the supply chain. Although, beyond the scope of this particular research, we will investigate the general effects of different supply chain designs on the selection of candidate geographical regions.

3.5 Phase I: Stochastic Programs for Region Identification

While the deterministic formulation of the region identification model provides a general idea of the profitability that could lie in the different regions, it does not account for the variability that exists in several of the parameters considered in this problem. In the real-world application, the main farmer would first make a decision regarding the regions to incorporate, the type of crops to plant, and the additional technologies/resources needed without truly knowing the outcome of random parameters that directly affect production. Once the first stage decisions have been made, the decision-maker would wait and observe the realization of random environmental parameters to make the set of second-stage decisions, such as the amount of resources to use, the amount of crops to harvest, and the best way to transport the harvest to the market place. Based on this basic set-up of this problem, we believe that the implementation of two-stage stochastic programs would be the best choice to improve upon optimal solutions found in the first part of the problem.

In a typical two-stage stochastic program, first stage variables are composed of those decisions made at the start of the planning period before the value of the random parameters ($\boldsymbol{\xi}$) are known. The second stage variables, also known as recourse decisions, are those made once the values of the random parameters are realized and known. Consequently, the objective function for these programs are constructed based on the deterministic values of the first stage decision variables ($\boldsymbol{x}$) and the set of second-stage decision variables ($\boldsymbol{y}$)

dependent on the expectation of the random parameters. This basic modeling structure is presented next:

$$\min c^T \boldsymbol{x} + Q(\boldsymbol{x})$$
$$s.t.\, A\boldsymbol{x} = \; b$$
$$\boldsymbol{x} \geq \; 0$$

where

$$Q(\boldsymbol{x}) = \; E_{\xi} Q(\boldsymbol{x}, \xi(\omega))$$

and

$$Q\big(\boldsymbol{x}, \xi(\omega)\big) = \; \min_{\mathbf{y}} \{q(\omega)^T \boldsymbol{y} | W\boldsymbol{y} = h(\omega) - T(\omega)\boldsymbol{x}, \boldsymbol{y} \geq 0\}$$

The calculation of $E_{\boldsymbol{\xi}} Q(\boldsymbol{x}, \xi(\omega))$ can be obtained by integrating over the random vector space $\boldsymbol{\xi}$ to obtain the expected value over different values for $\boldsymbol{x}$**.** However, this calculation can become tedious for higher dimensions of $\boldsymbol{\xi}$. Therefore, a common approach is to discretize its probability distribution function and estimate the discrete probability of each potential scenario.

As mentioned, we believe that the basic set-up of the two-stochastic programs fits the context of our problem well. As shown in Figure 3-8, our vision is that first stage variables would be comprised of the set of planting decisions, as well as technological investments, made at the beginning of a planting period under consideration. The second stage variables would then be the set of harvesting and resource allocation decisions based on the realization of random parameters, such as temperature and precipitation that also have an effect on yields, as well as well observed prices at the consumer market. Included in the second stage variables would be logistic decisions associated with transporting the product from the region to the market place. In this case, the objective would be to determine if there exists a particular planting period in which targeted technological investments would make it profitable to include a region into the production of a selected crop.

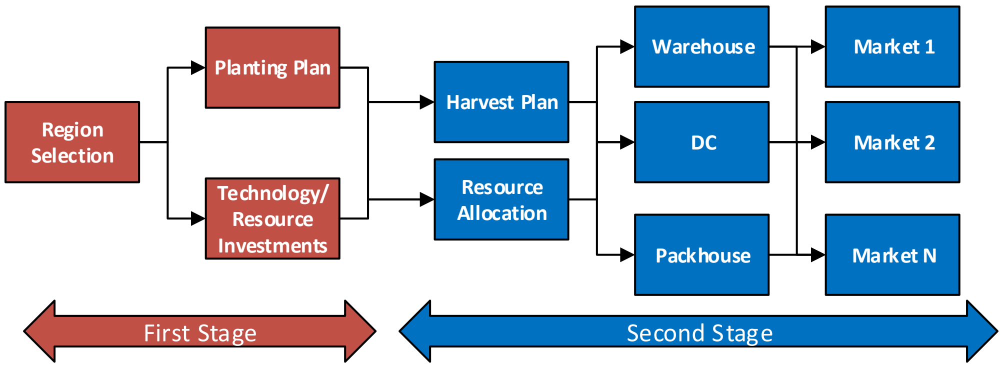


Figure 3-8: Two-Stage Stochastic Program

The decision to propose two-stage stochastic programs for addressing this problem is two-fold. Firstly, it provides the needed flexibility to incorporate variability from random parameters that is lacking in the deterministic version of this model. Secondly, the modeling structure is still manageable to handle a larger number of possible second stage scenarios, which is especially important under the context of this problem where there is more than one random parameter.

In similar fashion as for the deterministic model, second stage yields will depend on the decisions of the technological upgrades made. However, in this case, second-stage realizations of random yields may depend on the type of technologies used. For example, the range of potential yield outcomes may be reduced and more easily controlled by the implementation of irrigated greenhouse technologies regions where weather variability is more likely to impact production; however, this extra risk protection comes at a higher cost of implementation. Therefore, the model could further be used to perform sensitivity analysis on the tradeoff between the cost of implementation and the probability of different outcomes.

3.6 Phase II: Integrating Farmers in Selected Regions

While the outcome of the first phase consists of the selection of candidate geographical regions, alternative production options, and required technological infrastructure/ resources, the second phase of the study consists of an optimal allocation of technologies and resources to farmers *within* an identified region. In this case, the farmers will have to decide based on their own level of profitability whether to transition to a new set of higher value products. Similarly, the entity making the investment (referred to as to the central, decision-maker) would have to decide the physical assignment of space within static technologies (e.g. greenhouses, netting) such that the estimated production capabilities given by first phase results are met.

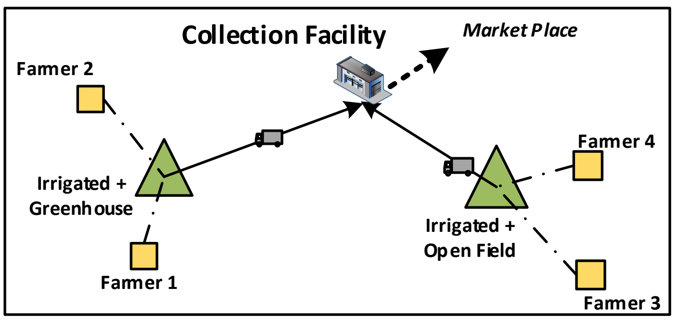


Figure 3-9: Decision Framework within Region

In this case, the main component of the second phase of the problem is the set of competing objectives between: (1) the maximization of profits received by the main farmer or the supply chain leader (i.e. investor of technologies and resources) and (2) the minimum level of profitability required by the set of individual growers in the region. This creates a sort of tradeoff scenario in which the profitability of the whole operation is dependent on the ability of the main farmer to assign technologies/resources such that demand is met and his profits are maximized, while also being able to offer competitive prices for crops, such that growers are incentivized to produce new set of products with the offered technologies.

To address this problem, we propose the use of mixed integer programs to construct the basic modeling structure. We first will use a contract selection structure in which we will design a set of offered contracts for the use of technologies at each potential location. In this case, each contract will have a defined crop price (paid by the main investor to the farmer) and production targets that farmers would have to meet using the installed technology at each location. From the set of possible contracts, the farmer would select the contract(s) that best fit his/her own profitability requirements. Each contract selection would also have an associated profit requirement values of the farmer that includes profit, cost, and any additional factors affecting his/her selection.

The objective function would be the maximization of profits from the perspective of the main farmer; he/she sells regional crop production at market price while paying the individual farmers the contracted price for their produced crops, as well as incurring the logistic costs associated with the technologies. An important constraint would be satisfying the minimum profit requirements set by the individual growers. In this set-up, all parties have to consider minimum profits and costs associated with different locations of the technologies and set of selected contracts. Included in these utilities/costs might be having to commute to the location of the static technology to produce the new set of crops (Figure 3-9).

3.7 Solution Approach

3.7.1 Phase I: Deterministic Model

One of the main difficulties in solving the proposed models (in both phases) is dealing with their large size and complexity. This is due to the large number of decision variables and constraints involved in the formulations. Fortunately, the structure of the model has characteristics that would allow the use of decomposition methods, which may greatly reduce convergence times. For this purpose, we propose the use of row and/or column generation methods, such as Benders and/or Dantzig-Wolfe, to break the overall problem into sub problems that are more efficiently solved. As part of this process, we plan to characterize the structures of these models in order to identify opportunities through which we can decompose them. (e.g. treating each region as independent subproblems).

Another implicit goal of this research is developing realistic and implementable tools. However, in order to achieve this, we will need to have access to accurate data. To do this, we will first attempt to determine the set of factors most important to crop yields and the effects that different technologies might have. We plan to talk to farmers, agronomists, and other people in the field, in order to determine the most important factors in order to reduce the amount of needed data. Furthermore, since it is highly likely that we will encounter situations in which we have missing values of data, especially for the most marginalized regions, we will search methods that can handle such situations. The search for this type of methods will be especially important when constructing probability distribution functions that can represent different second stage scenarios. As part of this process, we will also have to assess the effect of missing data on the solutions of the model.

3.7.2 Phase I: Stochastic Model

As part of this study, we also propose the use of two-stage stochastic programs to account for stochasticity in some of the parameters of the formulations. In this case, the inclusion of different scenarios into problem can greatly increase the complexity of the problem. However, as noted in Birge and Louveaux (1997), the structure of second-stage, discrete scenarios has the characteristics that would allow the use of decomposition methods, such as Benders (a row-generation technique), in order to reduce the time to convergence. In this part of the study, we plan to use row-generation techniques, such as L-shaped method. Another important component of stochastic programming is being able to generate discrete, second-stage scenarios. Generating these scenarios depends on the characterization of the sample space and probability distribution functions of the random parameters. In this case, focus should be given to the tradeoff between increased accuracy given by higher resolution on second-stage scenarios and the size of the solution space. Once we have defined a probability distribution function of random parameters, we plan to increase the resolution of second-stage scenarios space in an incremental manner. An alternative method that will be explored is the use of sampling procedures, such as those implemented in Santoso et al. (2005), who instead of completely enumerating second-stage scenarios, use sample average approximations to sample scenarios of the solution space. By using this kind of approach, we can considerably reduce the size of our problem.

3.7.3 Phase II: Allocation of Technologies and Resources

The basic premise of the second phase of the problem is being able to maximize the profits of the main farmer (SC leader) while simultaneously satisfying the minimum profit requirements obtained by the growers in the region. This is a slightly different modeling perspective as that from the first phase in which the decisions are made by a single, centralized decision-maker. In this case, the decision variables are the allocation of production contracts that dictate how technologies are shared. Since both the main farmer and the individual grower are searching for solutions that satisfy their own personal utilities, this leads towards solutions that are decentralized.

3.8 Validation and Case Study

From previous experience with working with growers in different crop markets, the industry is not accustomed to using very elaborate planning tools. Thus, the development of models that use as little information possible can be conducive to wider use of these kinds of tools. Also, the tools developed should give candidate solutions that are interpretable to farmers, so that they can be verified with information from real systems. Furthermore, we aim to develop models that are robust enough to handle different settings of the problem.

One should note that the number of variables in the model formulation is expected to be high and difficult to solve. However, we draw on examples from previous works, such as Ahumada et al. (2012) and, who use decomposition methods to break the overall problem into a series of subproblems as evidence that this type of problems can be efficiently solved. For example, in Ahumada et al. (2012), the authors use Bender's decomposition method, in similar fashion as proposed in this research, to create tactical planning models for the production and distribution of agricultural products under uncertainty. Similarly, Mason and Villalobos (2014) use Dantzig-Wolfe decomposition applied to iterative, auction-based mechanisms to coordinate seasonal production plans for a group farmers. We plan to use similar decomposition methods to break the overall problem in to subproblems that can be solved more efficiently.

One of the goals for the proposed models is usefulness and applicability to real settings. The ultimate test of the usefulness of our system is the demonstration that our model can identify candidate production regions, as well as provide a guideline on assigning technologies to growers in identified zones. (Mason and Villalobos, 2014) The models to be presented in the remainder of this dissertation have been subjected to a validation process through several case studies. These models can be separated into four main areas of research. The first is the development of an overall optimization framework that considers environmental conditions, market prices, plant physiological requirements, and logistic components in the identification of geographical regions with potential to produce high-value crops by combining yield assessment methods of fresh vegetables with supply chain planning (chapter 4). The second area is the implementation of stochastic decomposition methods that can handle a larger set of discretized scenarios created by variability introduced to the yield estimation methodology (chapter 5). The third is the inclusion of machine learning techniques to learn the relationship between first-stage solutions and generated yield scenario instances within a stochastic decomposition implementation (chapter 5). The fourth and final area is the development of a decentralized optimization framework that takes the solution from previous problem results to assign technology resource use and plan labor requirements within each region, and where farmers are assumed to be individual entities with their own minimum profitability requirement and an investor seeks to maximize the profitability of his/her investment (chapter 6).

4. DESIGN OF A COMPLEMENTARY FRESH-FOOD SYSTEMS

The development of a complementary fresh-food system is dependent on three main components. The first component is the ability to estimate yields of high-value, perishable crops as a function of few climate parameters. Since expected yields are estimated at a macro level, the availability of weather information across a large geographical area can limit the level of resolution for these estimates. However, the goal within the scope of this framework is to identify a preliminary set of candidate production strategies despite limited weather information. The second component of this framework is the development of an optimization model that inputs yield estimates at different time periods and outputs an investment, production, and market strategy that maximizes the profitability of a centralized decision-maker. Since horticultural production is highly dependent on random parameters (e.g. weather temperatures, precipitation, market prices), the third component is developing a framework flexible to consider variability within these variables. In this section, the first two framework components are introduced. First, an approximation methodology is developed that estimates yields as function of observed temperatures during different time periods. The second component is a deterministic optimization model defining the structure of the problem, and whose output is an optimal logistic strategy and combination of regions, crops, technologies, and markets that maximizes profit margins. The main output is the design of a supply chain based on complementary fresh food production based on regional weather patterns as shown in Figure 4-1.

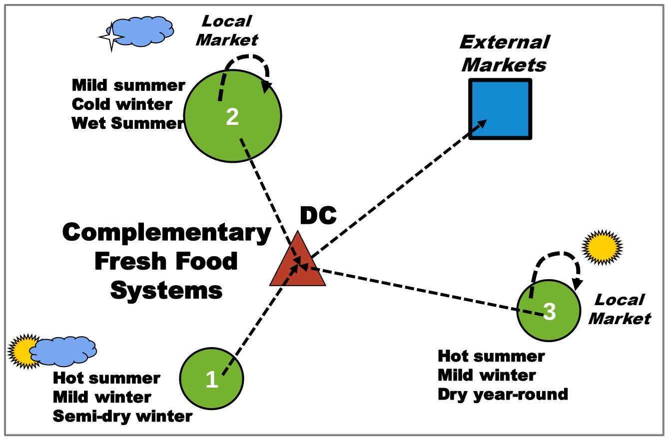


Figure 4-1: Design of Complementary Fresh Food Systems

The yield approximation framework developed in this section is a significant contribution to the agricultural planning literature. It takes an analytical and decision-maker perspective to estimating high-value crop yields, which differentiates it from previously developed tools that have been geared towards less perishable items, such as rice, beans, and corn. For example, previously developed tools do not consider the planting and harvesting patterns typical for highly-perishable items nor do they provide guidelines on optimal production periods. Therefore, this yield approximation framework adds a level of resolution to previous estimation methods. Most importantly, the framework is integrated with an optimization model whose decision-making component determines an optimal planting and harvesting plan. This facilitates the exploration of production areas that may have data collection capabilities of daily temperature and precipitation values but may not have historical production information. To the best of our knowledge, this work is the first to develop high-value agricultural planning models by incorporating yields directly as a function of complementary, environmental parameters. Within a practical implementation of this approach, the yield estimations and production strategies outputted by this framework would serve as a first-step before an in-depth assessment of region-specific characteristics, such as soil properties and land and labor availability.

4.1 Design for a Deterministic Framework

The general scheme of the deterministic framework is presented in Figure 4-2. The deterministic framework begins with the collection of historical data on daily precipitation and maximum and minimum temperature values for weather stations within a broad geographical area. The number of weather stations can be relatively high, especially as the scale of the implementation is increase (e.g. multiple U.S. states. Thus, the second step is to group these regions into homogenous clusters that share similar temperature and precipitation characteristics, which avoids redundancy in the model. In this manner, the complete weather station dataset can be reduced to a select few zones with distinct weather patterns but can be representative of the broader geographical area. The selection of this weather station subset could also be based on subjective criterion, such as their access to logistic infrastructure and proximity to urban centers. Once this regional subset has been selected, the next step is to translate temperature values within the distinct regions into yield estimates for different planting and technology decisions. These estimates would then be inputted into the optimization model as a parameter set. Similarly, price values in target wholesale markets would be collected, process, and inputted into the optimization model. Finally, one has the flexibility of constructing the general rules for the supply chain design of the operations.

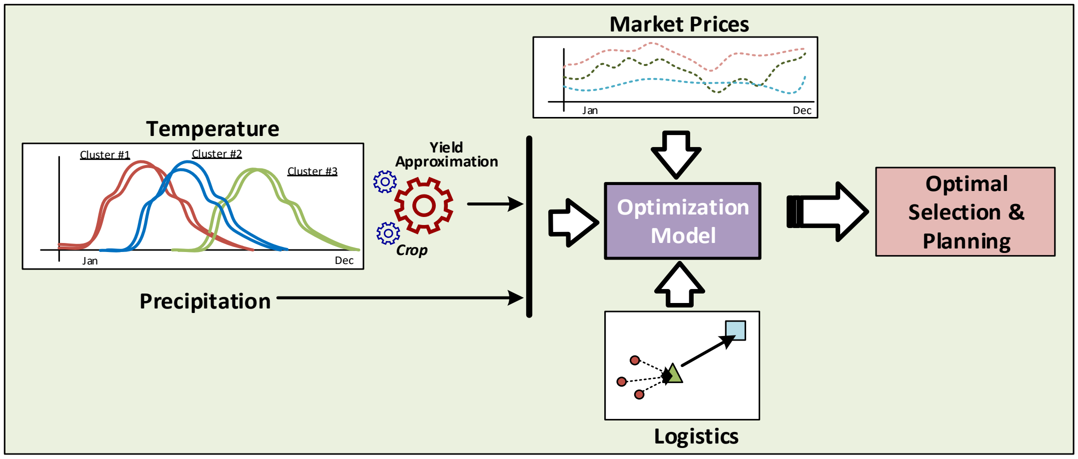


Figure 4-2: Design of Deterministic Framework

The output from the optimization model is a set of planting and technology investment decisions within each region. These planting decisions would include the type of crop planted for each technology at the most profitable period within each location. Based on these planting decisions, the optimization will also output the expected harvest amounts per week as a function of observed temperatures between the planting and harvesting periods. It also includes a selection of target wholesale markets for each crop by choosing the most profitable price pattern, as well as a set of consolidation points and modes of transportation for the constructed supply chain design. The framework is presented through a case-study applied to the U.S. Southwest within the states of Arizona and New Mexico.

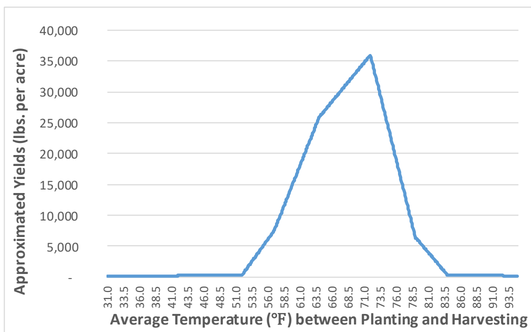
In the next few sections, the yield approximation framework is presented and applied to a set of selected crops. Then, a deterministic optimization model is developed that seeks to determine the optimal selection of crops, technologies, and zones that would maximize a centralized decision-maker's profitability. Finally, the section is concluded with a case-study applied to the U.S. Southwest.

4.2 Yield Approximation in Planning for Fruits and Vegetables

One of the most important production constraints of high-yield non-perennial, fresh vegetables is satisfying its physiological requirements. Thus, an important part of the proposed framework is developing an approximation method that can quantify yield estimates under varying environmental conditions, which can be used to answer basic implementation questions. For example, it may help answer whether tomatoes can be profitably grown during northern Arizona's cold winters using greenhouse technologies or whether nearby regions could complement local production and fulfill year-round offering. Furthermore, it allows the user to approximate yield patterns using different protective and enhancement technologies. To demonstrate the basic mechanism behind this framework, tomatoes, romaine lettuce, and bell peppers were selected given their known physiological requirements. Then, empirical information on actual production plans and harvesting schedules from currently producing systems in Culiacan (Sinaloa), México, are used to validate the basic concept behind this framework. To further validate these estimates, county level production from known producing regions, such as Yuma, AZ, are used to compare estimated yield per acreage outputs for romaine lettuce versus historical information.

The proposed yield estimation framework is comprised of two basic components. The first component is based on the total estimated yield per acre one would expect under different observed average temperatures. For example, the estimated marketable yield for an acre of planted tomatoes ranges from 16,000 to 36,000 lbs. (Stoddard et al., 2007). However, Adams et al. (2001) reported that a mean temperature of 64°F reduces tomato yield to 21%

of its optimal at 72°F, while temperatures of 79°F and 57°F reduced its yield to 18% and 75% of its optimal, respectively. Furthermore, average temperatures below 52°F (Criddle et al., 1997) and above 83°F showed no significant tomato yields (Vanthoor et al., 2011). Now, using this temperature information as breakpoints, a linear piecewise function is constructed to enable yield extrapolation for a range of average temperature values (Figure 4-3). With this function, one can obtain rough yield approximations of different vegetable items across distinct location-based, weather patterns.



Temp (°F)	Yield (lbs)
52	36
57	7,560
64	25,920
72	36,000
75	6,480
84	36

Figure 4-3: Tomato Yields as Stepwise Function of Average Temperature

Similarly, piecewise yield functions were used to approximate the outputs for the other vegetable items (Figure 4-2). For example, optimal temperature for romaine lettuce production is generally cooler, ranging from 55°F to 65°F (Drost, 2010), at which it can yield approximately 30,000 pounds per acre (USDA, 2015a). Yields under extreme temperatures below 40°F and above 85°F will result in poor or non-existent germination of lettuce seeds (Smith et al., 2003). On the other hand, bell pepper production is more adaptable to higher temperature. For example, under optimal conditions, bell pepper yields can reach approximately 36,600 lbs. per acre (Hartz et al., 2008), when average temperature range is approximately 73°F (Tewari, 2015). However, when temperature dip below 55°F or reach over 90°F, plant germination will be greatly affected (LeBoeuf, 2004). As part of these yield approximations, it was also assumed that if the maximum and minimum temperatures in a region surpass crops temperature thresholds, then yields would be reduced to 5% of their optimal.

Table 4-1: Estimated Yields under Varying Temperature

Bell Pepper		Romaine Lettuce	
Temp (°F)	**Yield (lbs)**	**Temp (°F)**	**Yield (lbs)**
55	1,830	40	1,567
73	36,660	60	37,200
95	1,830	85	1,567

The second component of the proposed yield estimation framework relates to the distribution of yields during a harvesting period. For example, based on previous conversations with tomato farmers from Sinaloa, if one plants a set of open-field acres during the first week of August (i.e. WEEK 1), the farmer expects to harvest during WEEK 13 through WEEK 28. During the first weeks, the farmer would harvest a small percentage of the total yield (e.g. 1%) gradually, increasing to 8% to 10% of the total, and back down to 1% by the end of the harvest period (solid line in Figure 4-4). Protective growing technologies may help extend the harvesting period by another four weeks by maintaining a controlled production environment, which in turn alters the distribution of the yield (dotted line in Figure 4-4). It can also be used to increase the obtained yields. Thus, this second component allows the user to design the harvest pattern that best matches the yield distribution given by empirical information and to provide rough extrapolation estimations when applied to other systems.

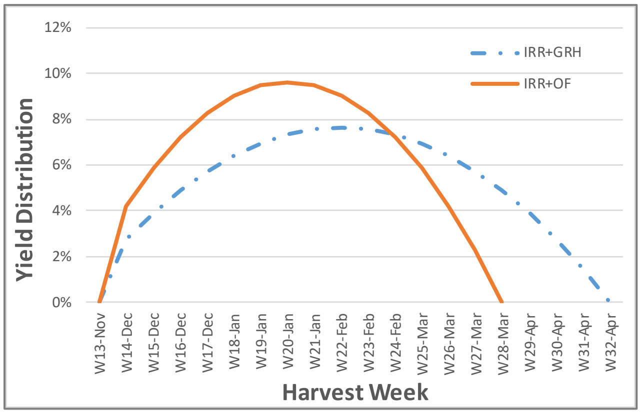


Figure 4-4: Distribution of Yields through Harvesting Period

By combining the information derived from both components, the farmer can then approximate expected yields based on the week the crop is planted and estimate how its harvest will be distributed during harvesting season. For example, in Figure 4-5, one can observe the expected tomato harvest in Culiacan, Mexico, by planting week (each line represents a different plant week). Based on these estimates, one could recommend planting tomatoes during the second week of September, given that the estimated harvest yields seem to be highest. However, one should note that this decision might change once the complete production schedule is aggregated with the production from other regions. Specifically, the optimization model would seek to schedule production as to maximize the combined profit margin.

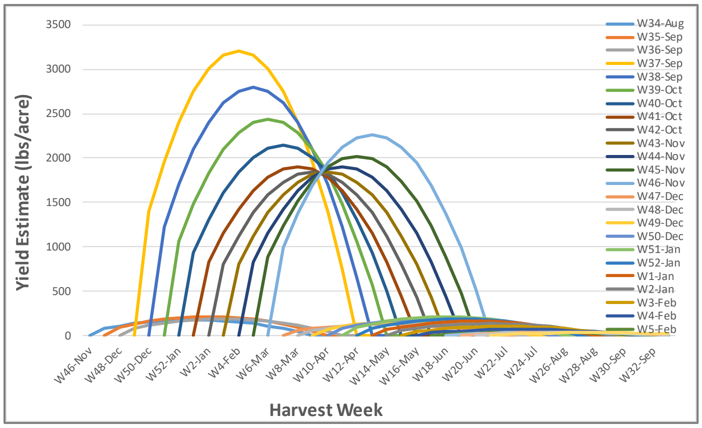


Figure 4-5: Tomato Harvests for Planting Weeks in Culiacan, Mexico

Collected empirical information was then used to test the validity of this approach for the set of fresh vegetable items outlined above (Ahumada, 2016). For example, in the case of fresh tomatoes and bell peppers, historical harvest projections and planning schedules from farmers in northwestern state of Sinaloa were used as a basis of comparison for the yield approximations from this framework. The tomato and bell pepper yield function detailed in Figure 4-3, above, was applied to historical temperature data to estimate yields as a function of observed average temperatures between different planting and harvesting weeks. Figure 4-6 provides a comparison between the farmer's actual harvest projections for an acre planted during the first week of September versus the framework estimation for tomato and bell pepper harvests. As one can observe, these estimations closely resemble actual projections as the average temperature in the region are close to the optimal during this period.

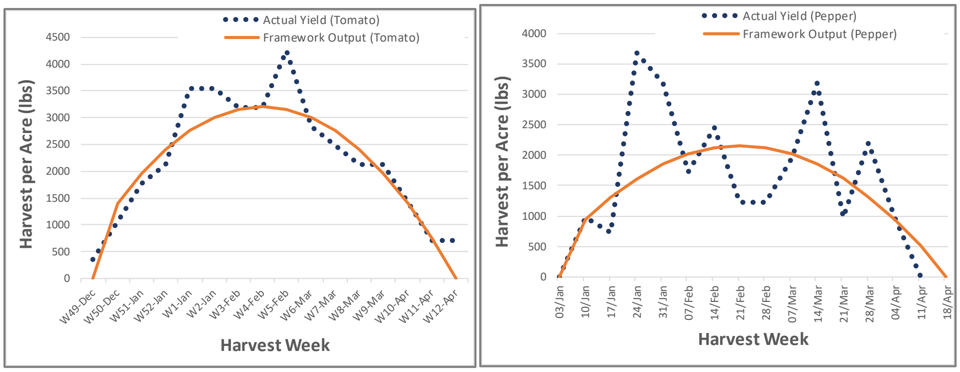


Figure 4-6: Framework Estimates versus Actual Yield Projections

A similar analysis was made to validate yield estimates for Romaine lettuce in Yuma, AZ, using historical production in the region. Information derived from the U.S. agricultural census shows that the yield per acre in this region is equivalent to approximately 36,000 lbs. (USDA, 2012), while the planting season typically begins anywhere between late September and mid-November (Wishon et al., 2015). With this information, one can compare the output from the yield approximation framework against historical values. In this case, Figure 4-7 presents the estimated harvest pattern provided by the model. In this figure, the y-axis is the harvested amount per week (lbs.), while the x-axis delineates a different harvest week. Moreover, each colored line in this figure represents the harvest pattern for each different plant week. The actual estimates for the first half of the planting season is detailed on the left of this figure. From this figure, one can observe that the planting week that resulted in the highest yield values were from late September to late October, which roughly matches the actual planting period of the region. Furthermore, the yield per acre estimated by the model was 36,500 lbs. for these two weeks roughly approximating actual values.

Week Planted	Yield (lbs.)
W34-Aug	32,899
W35-Sep	34,261
W36-Sep	35,283
W37-Sep	36,135
W38-Sep	36,497
W39-Oct	36,199
W40-Oct	36,199
W41-Oct	36,497
W42-Oct	36,305
W43-Nov	35,283
W44-Nov	34,261
W45-Nov	32,899
W46-Nov	31,366
W47-Dec	29,323
W48-Dec	27,279
W49-Dec	24,895
W50-Dec	22,511
W51-Jan	20,126

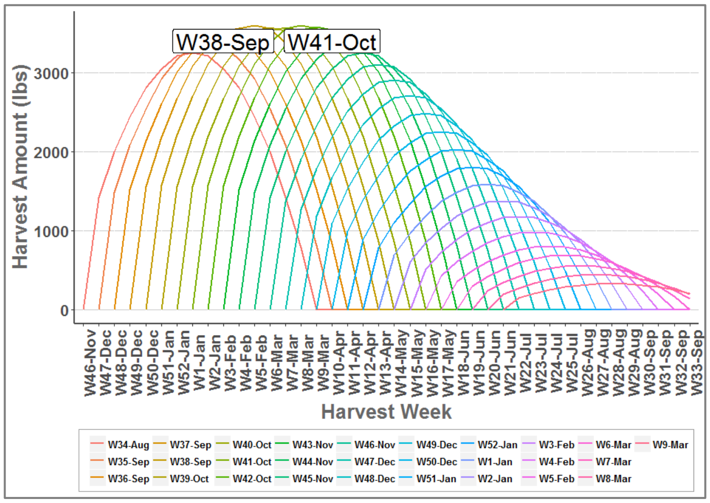


Figure 4-7: Estimated Harvest Pattern in Yuma, AZ

An additional, important physiological component considered in this framework are vegetable water requirements. In this case, the assumption is that these requirements will be met either through precipitation or additional water allocation to the region. Table 4-2 presents water requirements for each vegetable item and three irrigation systems, sprinkler, drip, and controlled. Even though the costs associated with a drip irrigation system are higher, it has been shown that irrigated vegetables require as much as 30 percent less water than sprinkler systems (Harrison, 2014). Based on available published reports, the amount of water required by tomatoes is 678,850 gallons per acre on average (Masabni, 2016), which translates to roughly 969,786 gallons under a sprinkler system. Similarly, Masabni (2016) estimates that lettuce and bell pepper require 271,540 and 814,620 gallons per acre, respectively, on a drip irrigation system from which open-field values can be calculated.

Under controlled environment, the efficiency of water efficiency is conservatively assumed to be ten times as efficient, when compared against that of open-field (Mattson, 2015).

Table 4-2: Water Requirements

Crop	Water Requirements (gal/acre) Controlled (control irrigation)	Protected (drip irrigation)	Open-Field (sprinkler)
Tomato	96,977	678,850	969,786
Lettuce	38,791	271,540	387,914
Bell Pepper	116,374	814,620	1,163,743

Overall, the major tradeoff of the proposed yield estimation framework is between simplicity and estimation accuracy. As mentioned, this framework allows the decision-maker to have an initial output estimate of a local fresh-food system under different temperature conditions. Once an initial set of regions have been identified, the second phase would be acquiring additional location-specific information, including land availability and temperature variations, within each individual zone, as well as specific planting strategies. Another advantage from this framework is that it allows the user to incorporate additional physiological components to this yield function, such as growing degree days and soil properties, to further refine these approximations. Finally, it is assumed that protective technologies can modify these estimates, by changing the average temperature conditions observed by the plant during production.

4.3 Supply Chain Modeling for Complementary Systems

Once a yield estimation framework has been developed, the second part is using this information within a decision-support tool. This section presents a deterministic, mixed-integer programming model that uses the set of physiological requirements from the previous section to construct optimal production planning models. Its scope is the strategic identification of potentially producing zones and development of tactical planning tools to coordinate production. The objective function is based on the profitability of investment, production, and logistic decisions made by a set of small, micro-farming systems located in distinct geographic regions. Among the set of parameters are the wholesale market prices and expected yields as a function of temperatures values. The model then determines the set of protective technologies that can modify the harvesting period and improve yields whenever profitable. Finally, this set-up would allow future incorporation of probabilistic behavior from environmental, yield, and market parameters.

The main objective of the mixed-integer program is to identify and coordinate production from a complementary set of fresh food systems. The main component for this model is determining when to plant and harvest the different crops within each region, as well as the type of protective technologies needed. The model is constrained by the amount of water available, the availability of land, and the amount of investment capital. For this model, the decision set is constrained to one year, in which long-term investment and operational costs are annualized. Finally, the mechanism behind this framework is exemplified through a case study that considers sets of farmers within different regions of the U.S. Southwest.

Sets:

$$
\begin{aligned}
z \in Z&: \quad \textit{Set of zones (regions) for production} \\
f \in F(z)&: \quad \textit{Set of farmers in zone } z \\
j \in J&: \quad \textit{Set of crops} \\
t \in T&: \quad \textit{Set of time periods} \\
c \in C&: \quad \textit{Set of customers/markets} \\
m \in M&: \quad \textit{Set of transportation models} \\
t_p \in T_p \subset T&: \quad \textit{Set of planting periods in } T \\
t_h \in T_h \subset T&: \quad \textit{Set of harvesting periods in } T
\end{aligned}
$$

Decision Variables:

B_{jfu}: $\begin{cases} 1 & \text{if technology } u \text{ is vailable to farmer } f \text{ ? } F(z) \text{ for crop } j \\ 0 & \text{otherwise} \end{cases}$

$X_{jfu}^{t_p}$: Yld of crop j by farmer f when planted at t_p using technology u

$MicroHarv_{jz}^{t_h}$: Amount of crop j harvested during t_h within zone z

$Pack_{jz}^{t_p,t_h}$: Amount of crop j packaged during t_h planted in t_p within zone z

$WA_z^{t_p,t_h}$: Additional water allocated to region z between t_p and t_h

$SLZ_{jfz}^{t_h,t}$: Qty. shipped of crop j from farmer m *to* region z at time t harvested at t_h

$SZD_{jzdm}^{t_h,t}$: Qty. shipped of crop j from region z *to* DC d at time t harvested at t_h on mode m

$SDC_{jdcm}^{t_h,t}$: Qty. shipped of crop j from DC d *to* market c at time t harvested at t_h on mode m

$\mathrm{Invw}_{\mathrm{jz}}^{\mathrm{t_h}}$: Inventory of crop j at zone z at time $\mathrm{t_h}$

$\mathrm{Invw}_{\mathrm{jz}}^{\mathrm{t_h}}$: Inventory of crop j at zone z at time $\mathrm{t_h}$

$AddWCap_z$ Additional warehouse capacity used at zone z

Parameters:

$Land_f$: Land available to farmer f

$Yld_{jzu}^{t_p}$: Yld of crop j in zone z when planted in t_p using technology u

$YDis_{jzu}^{t_p,t_h}$: Yld distribution when planted/harvested in t_p/t_h using technology $u$for crop j in region z

$\mathrm{LRainRec}_{\mathrm{z}}^{\mathrm{t_p,t_h}}$: Rain received between t_p and t_h in region z

$WReq_{ju}$: Water requirements for crop j using technology u

$maxl_j$: Maximum number of acres that can be planted of crop j

$minl_j$: Minimum number of acres that can be planted of crop j

$\mathrm{MaxDem}_{\mathrm{jc}}^{\mathrm{t}}$: Demand for crop j by customer c at time t

MPr_{jc}^t: Price offered for crop j by customer c at time t

$CropOper_j$: Maximum number of operations producing crop j

$Ctech_u$: Amortized investment cost of technology u

$Cplant_j$: Cost of planting a full acre of crop j

$Coper_u$: Cost of operating technology u for one year

$CTLZ_z$: Transportation cost from region z warehouse

$CTZD_{zd}$: Transportation cost from region z to DC d

$CTDC_{dcm}$: Transportation cost from DC d to customer c using transportation mode m

$Ccase_j$: Packaging cost for crop j

$Cwater_z$: Cost of additional water allocated to region z
$Cavail_f$: Capital available to farmer f
LT_{dc}: Transportation time between DC d and customer c

Maximize:

$$\begin{aligned}
&\sum_{jdmt_htc:t=t_h+LT_{dc}} SDC_{jdcm}^{t_h,t} * Mpr_{jc}^{t} \\
&- \sum_{jqzt_ht} Invw_{jz}^{t_ht} * Cw_z - \sum_{jqdt_ht} Invd_{jd}^{t_ht} * Cd_d \\
&- \sum_{t_pt_hz} WA_z^{t_pt_h} * Cwater_z - \sum_{z,t_h,j} Pack_{z,j}^{t_h} * Ccase_j \\
&- \sum_{jdmt_htc} SDC_{jdcm}^{t_h,t} * CTDC_{dcm} - \sum_{jzt_htd} SZD_{jzdm}^{t_h,t} * CTZD_{zd} \\
&- \sum_{jft_htz:f\in F(z)} SLZ_{jfz}^{t_h,t} * CTLZ_z \\
&-M * \sum_{z} AddWCap_z \\
&- \sum_{t_pjfu} X_{jfu}^{t_p} * CPlant_j - \sum_{t_pjfuz:f\in F(z)} X_{jfu}^{t_p} * [Ctech_{uz} + Coper_{uz}]
\end{aligned} \qquad \textbf{Eq. 4-1}$$

Subject to:

$$\sum_{juf:f\in F(z)} B_{jfu} * Ctech_{uz} \leq Cavail \qquad \textbf{Eq. 4-2}$$

$$\sum_{t_p} X_{jfu}^{t_p} \leq Land_f * B_{jfu} \qquad \forall f, j \in J, u \in U \qquad \textbf{Eq. 4-3}$$

$$\sum_{uf} B_{juf} \leq CropOper_j \qquad \forall j \qquad \textbf{Eq. 4-4}$$

$$X_{jfu}^{t_p} \leq maxl_j * B_{jfu} \qquad \forall t_p \in T_p, j, f, u \qquad \textbf{Eq. 4-5}$$

$$\sum_{t_pu} X_{jfu}^{t_p} \geq minl_j * \sum_{u} B_{jfu} \qquad \forall t_p, j, f \qquad \textbf{Eq. 4-6}$$

$$\mathrm{WA}_{\mathrm{z}}^{\mathrm{t_p,t_h}} \geq -\mathrm{LRainRec}_{\mathrm{z}}^{\mathrm{t_p,t_h}} + \sum_{\mathrm{jfu:f\in Z(f)}} \mathrm{WReq}_{\mathrm{ju}} * \mathrm{X}_{\mathrm{jfu}}^{\mathrm{t_p}} \qquad \forall t_p, t_h \in T_h,\ z \in Z \qquad \textbf{Eq. 4-7}$$

$$\mathrm{MicroHarv}_{jz}^{t_h} = \sum_{t_p} X_{jfu}^{t_p} * \mathrm{YDist}_{jzu}^{t_p,t_h} * \mathrm{Yield}_{jzu}^{t_p} \quad \forall t_h, j, z, u, \; f \in F(z) \qquad \textbf{Eq. 4-8}$$

$$\mathrm{SLZ}_{jfz}^{t_h,t} = \sum_{u:q=2, f \in F(z)} \mathrm{MicroHarv}_{jfu}^{t_h} \quad \forall t_h, j, z, \; f \in F(z) \qquad \textbf{Eq. 4-9}$$

$$\mathrm{Pack}_{jz}^{t_h} = \sum_{fu:q=2, f \in F(z)} \mathrm{MicroHarv}_{jfu}^{t_h} / \mathrm{PckgWeight}_j \quad \forall t_h, j, z \qquad \textbf{Eq. 4-10}$$

$$\sum_{jq:q=2} \mathrm{Pack}_{jz}^{t_h} \leq \mathrm{WZCap}_z + \mathrm{AddWCap}_z \quad \forall z \qquad \textbf{Eq. 4-11}$$

$$\mathrm{Invw}_{jz}^{t_h t_h} = \sum_{f \in F(z)} \mathrm{SLZ}_{jfz}^{t_h t_h} \quad \forall t_h, j, z \qquad \textbf{Eq. 4-12}$$

$$\mathrm{Invw}_{jz}^{t_h t} = \mathrm{Invw}_{jz}^{t_h t-1} + \sum_{f \in F(z)} \mathrm{SLZ}_{jfz}^{t_h t} - \sum_{d} \mathrm{SZD}_{jzdm}^{t_h,t} \quad \forall t_h, t, j, \; t > t_h \qquad \textbf{Eq. 4-13}$$

$$\mathrm{Invd}_{jd}^{t_h t_h} = \sum_{z} \mathrm{SZD}_{jzdm}^{t_h,t} - \sum_{cm} \mathrm{SDC}_{jdcm}^{t_h,t} \quad \forall t_h, j, d \qquad \textbf{Eq. 4-14}$$

$$\mathrm{Invd}_{jd}^{t_h t} = \mathrm{Invd}_{jd}^{t_h t-1} + \sum_{z} \mathrm{SZD}_{jzdm}^{t_h,t} - \sum_{cm} \mathrm{SDC}_{jdcm}^{t_h,t_h+LT_{dc}} \quad \forall t_h, t, j, d, \; t > t_h \qquad \textbf{Eq. 4-15}$$

$$\sum_{t_h zqmc: t_h + SL_j \geq t \geq t_h} \mathrm{SDC}_{jdcm}^{t_h,t_h+LT_{zc}} \leq \mathrm{MaxDem}_{jc}^{t} \quad \forall jct \qquad \textbf{Eq. 4-16}$$

$$\sum_{t_h zqmc: t_h + SL_j \geq t \geq t_h} \mathrm{SDC}_{jdcm}^{t_h,t_h+LT_{zc}} \geq \mathrm{MinDem}_{jc}^{t} \quad \forall jct \qquad \textbf{Eq. 4-17}$$

The objective function of this framework (Eq. 4-1) is to maximize overall profitability by considering investment, production (etc. planting, water), and logistic decisions under different market price behavior and outputted harvested yields. Constraint Eq. 4-2 dictates the availability of investment capital to allocate to famers. Constraint Eq. 4-3 limits each farmer's planted area by the amount of available acreage at each location. ConstraintEq. 4-4 Eq. 4-4 is a logic condition that limits the number of farmers that can produce each crop. Constraint Eq. 4-5 constraints that the total amount of acreage does not surpass the available land, while constraint Eq. 4-6 sets the lower bound on the number acres that will be planted if a crop and technology is selected. Constraint Eq. 4-7 assures that the crop's water requirements are met for those crops that are planted either through rainfall or additional allocated water. Constrain Eq. 4-8 sums the production from the individual farmers within each zone according the total expected yield and its distribution across the harvesting period based on the amount of acreage planted, technology used, and regional weather patterns. Constraint Eq. 4-9 moves the product from the production site to collection facilities located on-site. Constraint Eq. 4-10 keeps track of the number of packaged items in warehouses on-site, while constraint Eq. 4-11 assures that the capacity of these facilities is not surpassed. Constraints Eq. 4-12 and Eq. 4-13 keep track of the shipment and inventory at these facilities. Note that all shipments are sent to a distribution center, handled by constraints Eq. 4-14 and Eq. 4-15, from which they are sent to wholesale markets. Finally, constraints Eq. 4-16 and Eq. 4-17 assure that the amount sent to the market is not larger than the actual harvested production, as well as assures that it does not violate maximum demand restrictions.

The output from the model consists of optimal production schedules, technological investments, and shipment quantities to wholesale markets. One should note that the supply chain of the optimization forces production from each zone be sent to distribution centers in major urban centers. The ultimate objective is to maximize the profitability of a centralized, local agricultural system based on small-scale operations. Note that in section 6, the problem is analyzed through a decentralized, decision-maker viewpoint in which an investor and individual farmers have individual minimum profit requirements. Finally, it should be noted that the optimization is solved using CPLEX 12.5.0 optimization suite and coded in A Mathematical Programming Language (AMPL). The modeling and optimization are implemented on an Intel Core i7-6700, 3.40GHZ computer with 16.0 GB of memory.

4.4 Case Study in U.S. Southwest in Arizona and New Mexico

To demonstrate the functionality of this modeling framework, an exemplary case-study is applied to various locations within the states of Arizona and New Mexico. Among the regions selected are the metropolitan areas of Phoenix, AZ, and Albuquerque, NM, due to their high urban population and contrasting temperature and precipitation traits. This case study includes the region surrounding Yuma, AZ, as it is the highest agricultural producing region in both states and is a major U.S. producer of romaine lettuce during the winter season (Wishon et al., 2015). Additional regions considered in this case-study are those from locations with distinct weather patterns but that also have access to logistic infrastructure and urban centers. The modeling framework is used to estimate the production potential for three vegetable items, bell pepper, romaine lettuce, and tomato (plum type). These items were selected based on their relatively high commercial value and the fact that they are widely consumed. For example, tomato is the second highest consumed vegetable in the U.S. at 17.3 pounds per capita, romaine lettuce is fourth with 10.6 pounds per capita, and bell pepper is fifth with 9.2 (onions and head/iceberg lettuce are first and third with 17.5 and 13.1 pounds per capita, respectively) (USDA, 2013). Meanwhile, the average U.S. *retail* prices for the latest available estimated year for tomato, romaine lettuce (full head), and green/red bell pepper are \$1.24, \$1.84, and \$1.84 per pound, respectively (USDA, 2015b). Not that in this initial case study, no differentiation is made between green and red bell pepper. In the next section, a sensitivity analysis is performed that specifically considers the red variety of the bell pepper, which in general has a higher market price.

4.4.1 Data Processing

Historical daily maximum, minimum, and precipitation datasets were collected from 1987 to 2016 for all available weather stations across the two states available from NOAA (2016a). Given that not all of the weather stations have a complete data set for this date range, only those stations having an availability of at least 70% were selected (i.e. the number of days with information divided by the total number of days). To further explore the relationships of weekly temperature and rainfall behavior between all weather stations, hierarchical clustering techniques are used to group these stations into homogenous groupings. The number of cluster groupings is determined by estimating the distance between average weekly precipitation and maximum and minimum temperature of the collected weather stations.

Table 4-3: Weather Station Weekly Temperature/Precipitation Data Structure

Station	Max Temp (Wk 1)	Min Temp (Wk 1)	Precipitation (Wk 1)	…	Precipitation (Wk 52)
1	$TMax_1^1$	$TMin_1^1$	$Prcp_1^1$	…	$Prcp_{52}^1$
2	$TMax_1^2$	$TMin_1^2$	$Prcp_1^2$	…	$Prcp_{52}^2$
…	…	…		…	
N	$TMax_1^N$	$TMax_1^N$	$Prcp_1^N$	…	$Prcp_{52}^N$

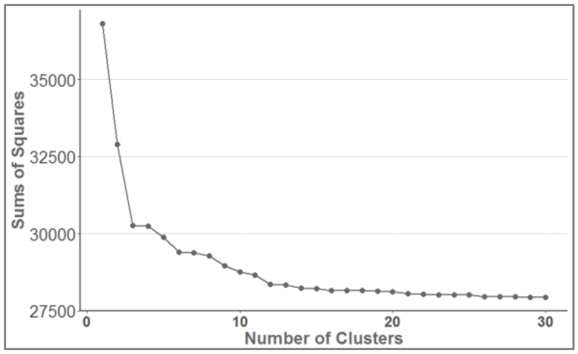


Figure 4-8: Sum of Squares Residuals vs. Number of Clusters

The data set structure used to perform clustering is presented in Table 4-3. The sum of squares distance between the cluster centroids and the stations within each grouping is calculated for different number of clusters. From Figure 4-8 one can observe that the value of the sum of squares levels off after 10 cluster groupings. From these 10 groupings, stations were arbitrarily selected from each cluster groupings based on their proximity to urban centers and access to interstate highways as shown in Figure 4-9. In this figure, the color and number of each point represents the grouping to which each station belongs. This means that all weather station with the same colored group have similar weekly temperature and precipitation traits through the course of one year. As one can observe, the initial region selection, Albuquerque, NM, Yuma, AZ, and Phoenix, AZ, belong to the first and second cluster groupings, respectively, while those with a label are additional locations considered within each group.

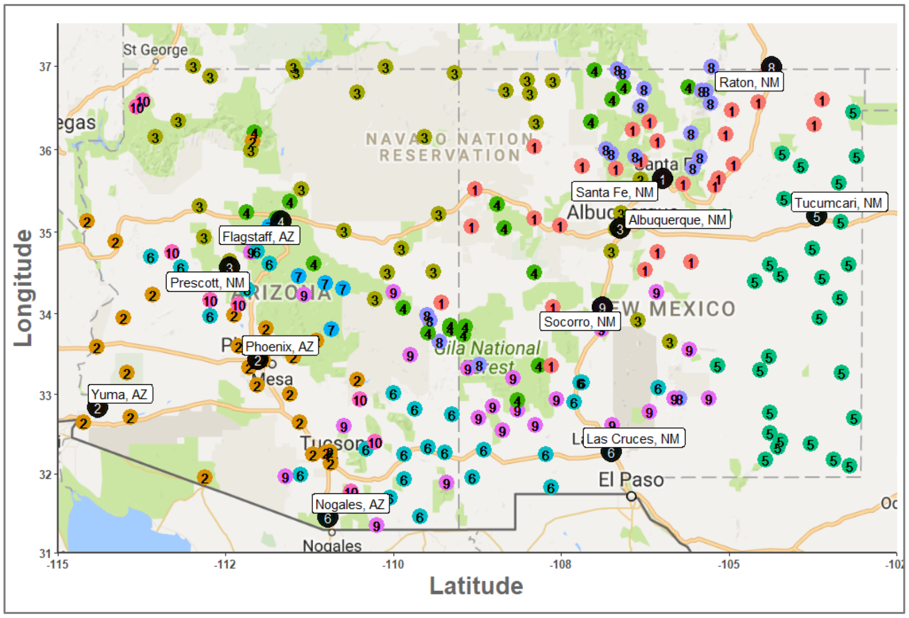


Figure 4-9: Cluster Groupings within Arizona and New Mexico

By looking closely to the weekly temperature and precipitation values within each cluster in Figure 4-10, one can observe the general characteristics of the different groups. In this figure, the top graph presents the weekly maximum and minimum temperature of the different stations that belong to each cluster. For example, cluster 2, to which Phoenix, AZ, and Yuma, AZ, belong, maximum and minimum temperatures are much higher, while the precipitation values are generally lower. Conversely, clusters 1 and 4 represented by Flagstaff, AZ, and Santa Fe, NM, have lower temperatures with much higher precipitation values. Cluster 8, represented by Raton, NM, seems to have the highest precipitation values and the lowest temperatures when compared to the rest. Additional cluster groupings represented in this case study are clusters 5, 6, and 9, which are represented by weather stations in Tucumcari, NM, Nogales, AZ, and Socorro, NM, respectively.

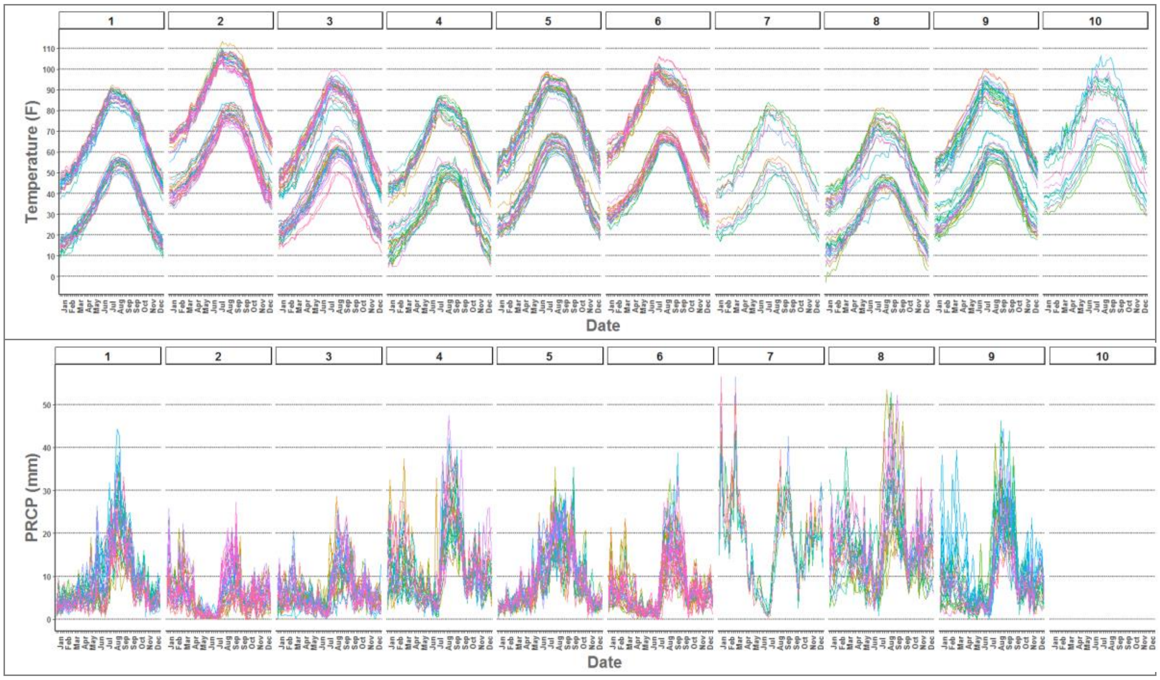


Figure 4-10: Weekly Environmental Behavior by Cluster Groupings

Weekly historical wholesale market prices were also collected for the three crops based on their years of availability, from 1998 to 2016. One should note that for tomato, only plum type (Roma) variety was selected, while for bell pepper and romaine lettuce there were no selection restrictions. This means that for bell pepper, all color varieties were considered, including red, green, and yellow colorings, while for romaine lettuce, hearts and crown cut leaves were included. This assumption is made since the output from these farming operations is considered relatively small and the marginal cost of post-harvest activities is also low. Nonetheless, in Section 4.6, an analysis is made to assess the impact that other variety selections would have on decision-making.

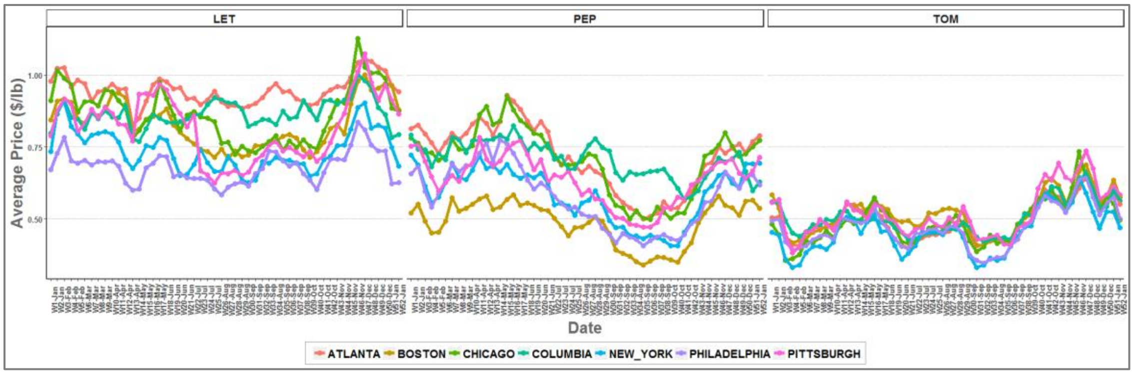


Figure 4-11: Average Weekly Market Prices

From the wholesale market prices presented in Figure 4-11, one can observe that peppers and romaine lettuce tend to have higher weekly prices when compared to tomato. In general, prices tended to be higher for the period between mid-March to early June, which may be due in part to the end of production seasons in Yuma and northern Mexico and the start of production elsewhere in the U.S. One should note that since these prices represent weekly averages, market prices are considered static within the deterministic optimization framework. This allowed the development of a more robust strategy that highlights the yield estimation mechanism behind a modeling framework directly dependent on environmental variables. Section 0 incorporates the variability within both weekly environmental parameters and weekly market prices.

4.4.2 Estimated Costs per Technology, Zone, and Crop

Another component of the optimization framework is the parameter set defining the optimization model. Estimated costs for each technology and each region are presented in Table 4-4 and Table 4-5, respectively. In this case, the estimated up-front, investment cost for an acre under a fully controlled environment is estimated to be approximately $1 million, while the annual energy cost is $125,000 (Mattson, 2015). Similarly, the investment cost for an acre equipped with greenhouse and irrigation technology was estimated to be approximately $502,144 (Zhang, 2016). On a 10-year horizon and 7% interest rate, these values translate to amortized investment values of $255,346 and $71,494, respectively. Similarly, the upfront cost for an open-field installment is approximately $8,000 (Orzolek et al., 2016), which translates to an amortized value of $1,139 on the same 10-year horizon. The last column in Table 4-4 contains the operational costs for these three technologies. The estimated operational costs for greenhouse and open-field is approximated to be $336,129 (Zhang, 2016) and $5,611 (Orzolek et al., 2016). The estimated operational costs for a controlled environment is estimated to be the greenhouse value plus the annual energy cost given by Mattson (2015), which amounts to $461,000. Furthermore, it is assumed that amount available for up-front investment is $1 million to invest in the collection of regions within the two states. The final column in Table 4-4 are the estimated yield increase by type of technology.

Table 4-4: Cost Estimates (per acre) for Technologies

Technology	Amortized Investment	Operational Costs	Yield Increase
Controlled	$ 255,346	$461,000	10**x**
Protected	$ 71,494	$136,129	4**x**
Open Field	$ 21,461	$5,611	1**x**

Table 4-5: Water Cost Estimates per Region

Zone	Cost ($/gal)
Albuquerque, NM	5.37E-03
Phoenix, AZ	4.33E-03
Yuma, AZ	2.13E-03
Nogales, AZ	4.20E-03
Flagstaff, AZ	6.08E-03
Prescott, AZ	7.49E-03
Las Cruces, NM	3.68E-03
Socorro, NM	3.64E-03
Santa Fe, NM	7.50E-03
Raton, NM	4.64E-03
Tucumcari, NM	5.98E-03

Estimated water costs per gallon are given in Table 4-5. Water rate estimates for commercial use are based on surveys performed in the states of Arizona and New Mexico by WIFA (2015) and NMED (2016), respectively,. From this table, one can observe the degree of variations of water costs between the regions within each cluster. In Arizona, the lowest water rates were estimated within Yuma and Douglas, AZ, while in New Mexico, Socorro and Las Cruces observed the lowest rates. The highest surveyed rates were observed in Santa Fe, NM, and Flagstaff, AZ. As it is shown in the next section, variations within regional water costs may impact investment and planting decisions.

Table 4-6: Planting Costs per Crop

	Planting ($/acre)
Tomato	$6,750
Lettuce	$8,500
Bell Pepper	$9,500

Costs for planting an acre of each crop type are estimated and summarized in Table 4-6. The estimated cost per acre for tomato production is approximately $6,750 (Orzolek et al., 2016), while for romaine lettuce production, the estimated cost is $8,500 (Tourte et al., 2015). Finally, estimated costs for bell pepper production is approximately $9,500 (Kaiser and Ernst, 2014). Additional logistic costs considered in the case study are summarized in Table 4-7, which include transportation and inventory costs while moving the product from its point of production to the consumption market.

Table 4-7: Logistic Costs

	Cost	Unit
Transportation	$0.01(truck) /$0.50 (air)	$/lb./mile
Inventory at Zone	$0.05	$/lb.
Inventory at DC	$0.01	$/lb.
Package Cost	$0.50	$/Package

Additional assumptions made for the case study implementation were that all regions would be comprised of one location, each with a 2-acre land capacity. It is assumed that the number of acres planted during any given week is less than 0.5 acres, while the minimum number of acres planted per crop is 0.1. In terms of the logistic strategy, it is assumed that all production will be sent from the zone to a distribution center located within a major urban center before reaching end consumers. Lastly, it is also assumed that the investor would have $1.5 million in up-front capital to invest in the development of the fresh-food production system.

4.5 Case Study Framework Output

This section presents the output from the deterministic framework demonstrating its functionality when exploring potential fresh food system. The output details the optimal planting and harvesting strategy for a select set of regions, crops, technologies, and markets, as well as a restricted investment budget. The planting strategy can be defined as an optimal schedule for each region, crop, and technology across the different weeks of the year, while the harvesting strategy defines the period through which each crop will be harvested. Furthermore, a logistic strategy is also drawn that details how the crop should move from its source of production to the end consumers. As mentioned, the objective is to determine the optimal decision combination that maximizes overall profit margins, from which one can analyze the revenue and cost aspects of different strategies. Finally, sensitivity analyses can be used to explore additional opportunities within the operations by focusing on specific crop varieties.

The planting strategy is comprised of the set of technologies and crops to plant within each region based in part on environmental and market patterns. An example from this output can be observed in Table 4-8, which shows the total number of acres planted in each region by technology and crop. For example, the optimization output suggests a total of 5 acres planted of Romaine lettuce in the regions of Socorro, Las Cruces, and Yuma under open-field conditions. The output also suggests small complementary production in Nogales and Phoenix that would be planted and harvested at different points of the season.

Table 4-8: Acreage Planted in Zone per Selected Technology and Crop

	Controlled			Protected			Open-Field		
Zone	**Let**	**Pep**	**Tom**	**Let**	**Pep**	**Tom**	**Let**	**Pep**	**Tom**
Albuquerque, NM	0	0	0	0	0	0	0	0	0
Flagstaff, AZ	0	0	0	0	0	0	0	0	0
Raton, NM	0	0	0	0	0	0	0	0	0
Nogales, AZ	0	0	0	0	0	0	1.2	0	0
Phoenix, AZ	0	0	0	0	0	0	0.7	0	0
Prescott, NM	0	0	0	0	0	0	0	0	0
Santa Fe, NM	0	0	0	0	0	0	0	0	0
Socorro, NM	0	0	0	0	0	0	5	0	0
Las Cruces, NM	0	0	0	0	0	0	4.7	0	0
Tucumcari, NM	0	0	0	0	0	0	0	0	0
Yuma, AZ	0	0	0	0	0	0	5	0	0

Figure 4-12 presents a bar plot of the planting schedule suggested for each region and crop. In this figure, since only one crop is suggested for planting there is only a single level. The x-axis represents the different planting weeks, while the height of each bar equals the number of acres planted by region. For example, one can observe that planting lettuce under open-field conditions in Yuma, AZ, is suggested for the months of October and early November. Conversely, open-field lettuce planting is suggested for the spring season in cooler climate regions, such as Socorro and Las Cruces. Small amount of lettuce planting is suggested at different times throughout the planting season in Nogales and Phoenix to complement main production. Finally, one can observe that tomato and bell pepper planting within each region is not selected within any of the regions.

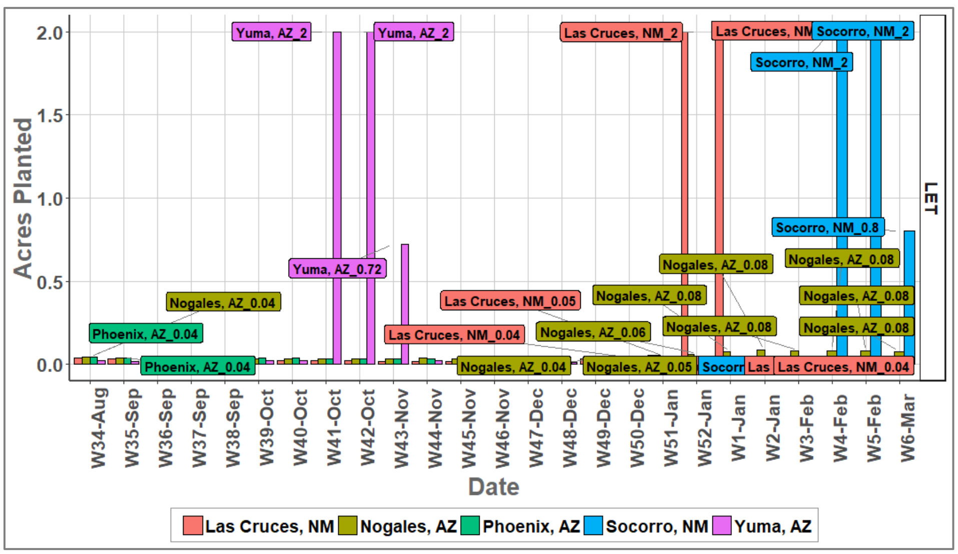


Figure 4-12: Number of Acres Planted per Zone and Technology by Week

Based on the number of acres planted, one can determine optimal harvest patterns depending on the crops selected (Figure 4-13). For example, lettuce harvest from warm regions such as Yuma can be expected from mid-December to mid-May, while complementary production from Las Cruces and Socorro is observed later in the harvesting season. Similarly, a small complementary system can be observed between Nogales and Phoenix that is small in comparison. One should note the complementary characteristics of the planting and harvesting period by seeking to fulfill yearly offering to improve overall revenues.

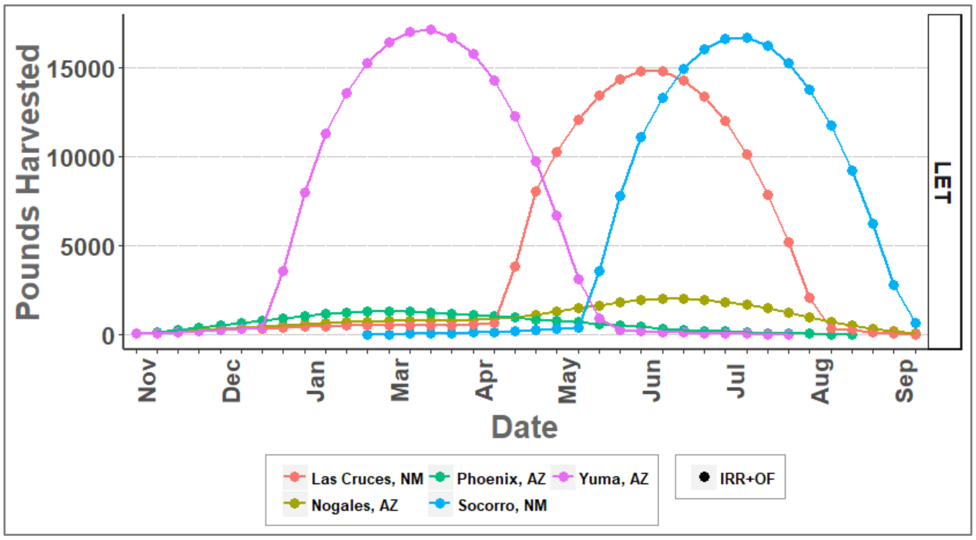


Figure 4-13: Weekly Harvest Quantities per Geographic Zone and Crop

From the framework output, one can also delineate a potential logistic strategy. In the left-hand side of Figure 4-14, the shipments from the individual regions to distribution centers can be observed for Albuquerque, Phoenix, and Tucson. This figure is segmented by crop and distribution center combinations where the x-axis dictates the shipment week, while the y-axis represents the crop quantity sent from each region to distribution center. From this figure, one can observe that consolidating production in Albuquerque and Phoenix seem like a better option, based on its high product movement, while Tucson is dedicated to consolidating shipments from the Nogales region. On the right-hand side of this figure, one can observe the second-leg of the logistic shipments from the DC's to the wholesale markets. Each colored point represents the DC origin for each weekly shipment, while the shape represents the mode of transportation. From this figure, one can observe three main wholesale markets using truck as the main mode of transportation, where lettuce production seem most profitable in the Atlanta and Chicago areas. Shipments to Chicago is sporadic and comes from distribution centers in Albuquerque, Phoenix and Tucson.

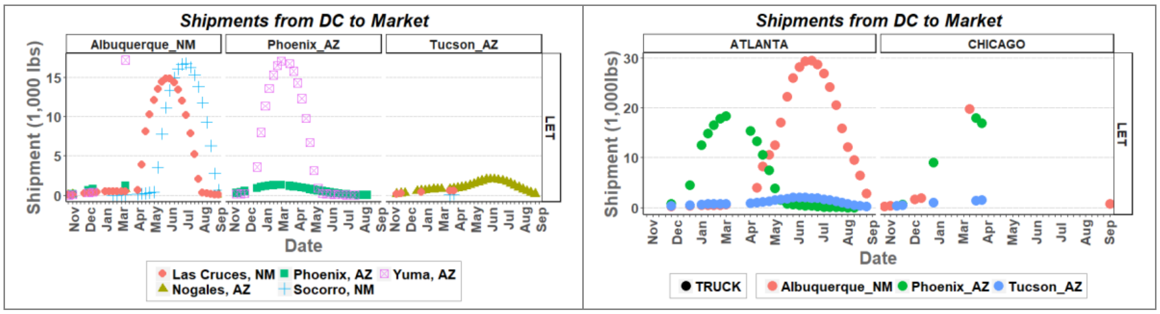


Figure 4-14: Shipment Strategies from Zones to Market

Finally, based on this planting, harvesting, and shipment schedule, an estimate is given of the expected revenues and costs incurred by the different participating regions. Figure 4-15 presents the estimated costs by zone and crop for the selected production strategy. Each color in this figure describes a different type of cost. From this figure, one can observe the cost segmentation for the different regions and crops, in which investments are amortized on a 10-year horizon at a 7% rate. For example, production for lettuce is dictated mostly by water costs, while the amortized investment and planting costs are relatively lower. One should not that the case in which revenues merit investments in higher technology options (e.g. protected or controlled), investment costs would be the driving cost.

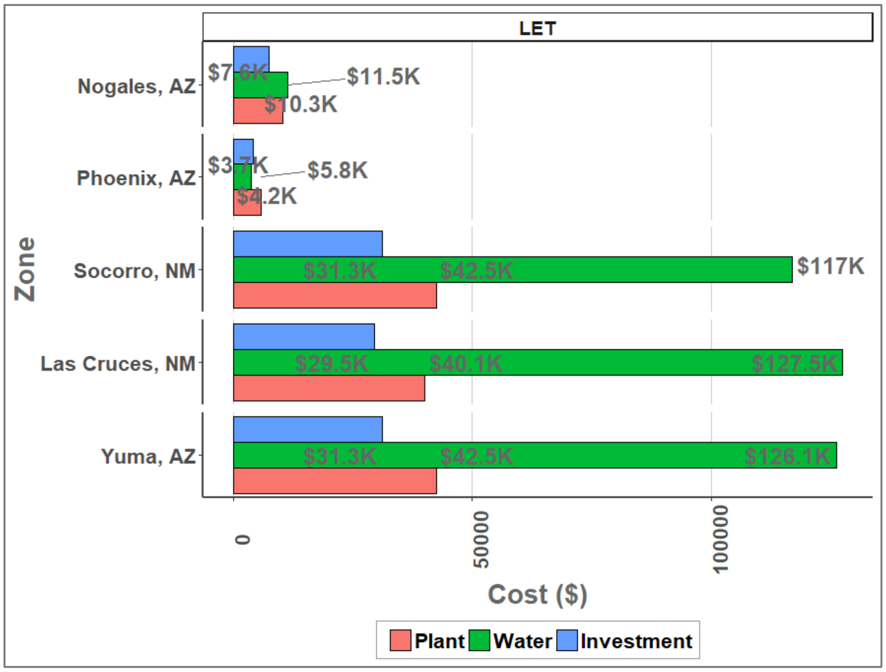


Figure 4-15: Estimated Costs by Type and Zone

Figure 4-16 presents the expected revenues per market and crop. Each level is a different wholesale market. One can observe markets in Chicago and Atlanta provide relatively high annual revenues. Overall, the most lucrative market place seems to be Atlanta with revenues over $500,000, while the Chicago market is estimated to generate approximately $80,000. Next, an analysis is performed to gauge how this would translate to actual expected profits and rates of return.

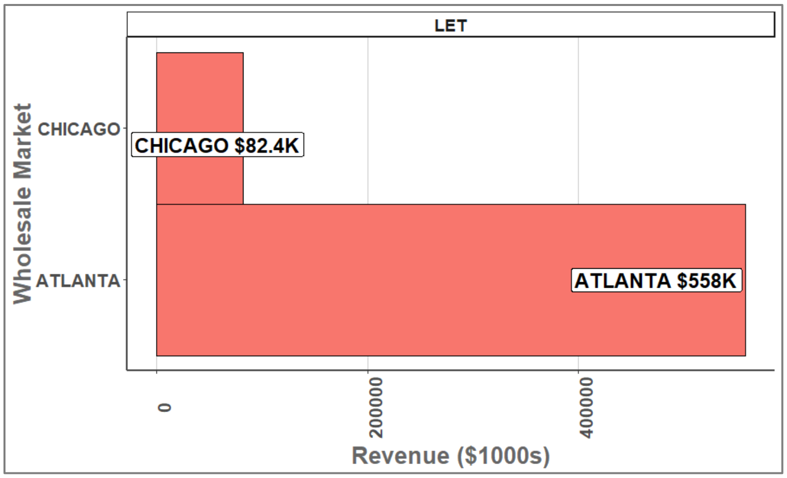


Figure 4-16: Revenue by Market and Crop

The last item in this profitability assessment are the revenues, costs, and rate of returns for each of the crops, as shown in Table 4-9. In this table, the first column presents in the total amount of shipments (in 1000 lbs.) made to consumer market, while the second column represents the revenues from these shipments. The second and third column represent the production (planting and water) and logistic costs for each crop, while the fourth column is the aggregated investment. The last column in this table is the calculated rate of return on investment for each of the crops. The rate of return (ROR) was calculated using equation Eq. 4-18.

$$ROR = \frac{(Revenues - Production - Logistic) - Investment}{Investment} \qquad \textbf{Eq. 4-18}$$

From this table, one can observe that lettuce observes the highest aggregated revenue compared to the rest of the crops, while also incurring higher production, logistic, and investment costs. The expected rate of return is high given that the level of investment from open-field operations comparatively low to the amount of expected revenues. From this collection of options, one can argue that lettuce production is a profitable opportunity given its high expected revenue stream and acceptable rate of return.

Table 4-9: Rate of Return on Investment by Crop

Crop	Shipments (1000 lbs.)	Revenues ($1000s)	Costs ($1000s)			Ann.
			Production	Logistic	Investment	ROR
LET	582	$640	$527	$104	$5	**5%**
PEP	-	-	-	-	-	**-**
TOM	-	-	-	-	-	**-**

As noted earlier, the development of the framework allows the user to explore different options. In this case, it seems that warm weather products, such as tomatoes and bell peppers, are relatively not attractive for the climate patterns in the region. Nonetheless, by focusing on specific varieties of this crop, one can potentially identify opportunities for these products. In fact, the next section analyzes the profitability from engaging in this specific variety of bell pepper.

4.6 Sensitivity Analysis

One of the benefits from this framework is that it allows the user to specify varieties for each vegetable. For example, one of the initial assumptions made for our case study is that the farmer may choose to produce both green and red bell pepper varieties. As a result, no distinction is made for weekly average market prices between the two varieties. However, if one considers specific varieties that command relatively higher market prices ($1.41/lb. for fresh green pepper versus $2.28/lb. for red pepper according to 2013 estimations (USDA, 2015)), then one can estimate how this would affect production and harvesting, as well as overall profitability of the operations. This section details the output from the framework when focus is given specifically to red variations of bell peppers.

Table 4-10: Acreage Planted in Zone per Selected Technology and Crop Varieties

	Controlled			Protected			Open-Field		
Zone	**Let**	**Pep**	**Tom**	**Let**	**Pep**	**Tom**	**Let**	**Pep**	**Tom**
Albuquerque, NM	0	0	0	0	5	0	0	0	0
Flagstaff, AZ	0	0	0	0	0	0	0	0	0
Raton, NM	0	5	0	0	0	0	0	0	0
Nogales, AZ	0	0	0	0	5	0	0	0	0
Phoenix, AZ	0	0	0	0	0	0	0.7	0	0
Prescott, NM	0	0	0	0	0	0	0.6	0	0
Santa Fe, NM	0	0	0	0	0	0	0.5	0	0
Socorro, NM	0	0	0	0	5	0	0	0	0
Las Cruces, NM	0	0	0	0	0	0	4.7	0	0
Tucumcari, NM	0	0	0	0	0	0	1.0	0	0
Yuma, AZ	0	0	0	0	5	0	0	0	0

By concentrating on the red bell pepper variety, the planting strategy for each crop and region is changed. Given the higher profitability in the red pepper variety, a higher focus is given towards high-yield production. For example, one can observe in Table 4-10 that protected technology operations are suggested for red pepper varieties in Albuquerque, Raton, Nogales, Socorro, and Yuma. In this product variety scenario, less consideration is given to lettuce production, in which Las Cruces and Tucumcari expect the highest planting quantities. Figure 4-17 details the planting (top) and harvesting (bottom) schedule for the red bell pepper variety based on this production strategy. One can observe similar complementary production patterns among the regions. However, higher activity is observed for bell pepper production. One should note production in Yuma has switched from lettuce to pepper. It is also important to note the switch to protective and controlled technologies. In the bottom of Figure 4-17, harvesting patters for these planting decisions observe year-round lettuce and pepper offering.

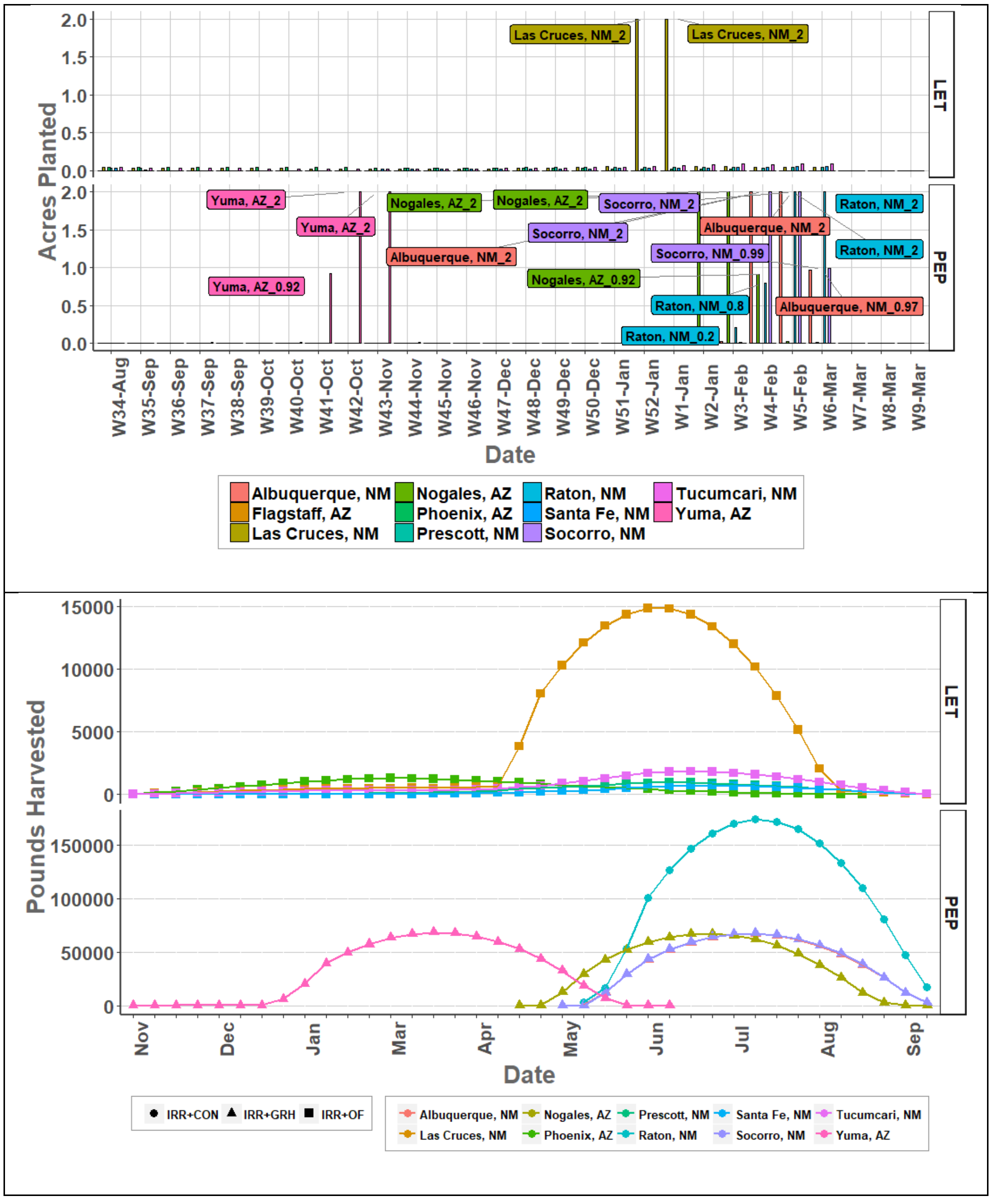


Figure 4-17: Weekly Plant/Harvest Schedule (w/ Bell Pepper: Red Variety)

In general, the logistic strategy remains similar to the previous to the undifferentiated option. As shown by Figure 4-18, Albuquerque remains the key distribution center for the three production operations. Phoenix and Tucson receive mostly dedicated production from Nogales, Prescott, and Yuma. It is interesting to note that Phoenix would consolidate production during the beginning of the harvesting season, while Albuquerque and Tucson serve to consolidate production later in the season (top of Figure 4-18). In the bottom of this figure one can observe that Columbia and Atlanta remain the target wholesale markets. However, shipments to Chicago and Pittsburgh can be observed intermittently throughout the harvesting season.

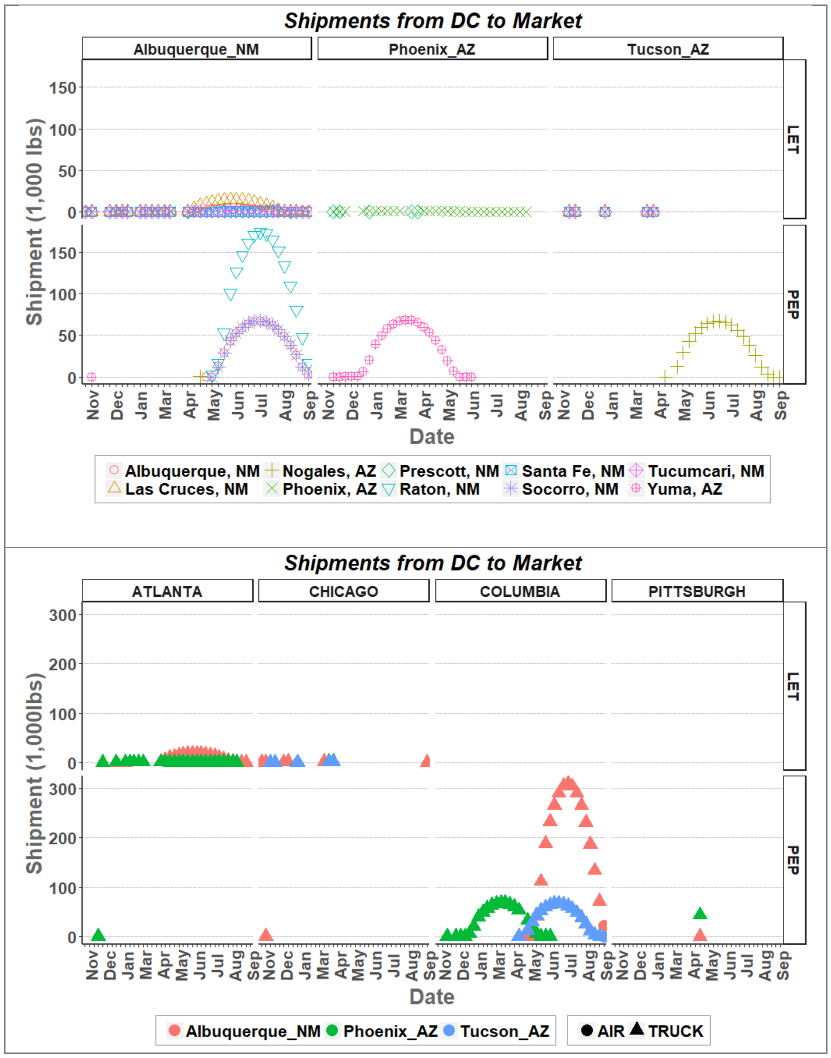


Figure 4-18: Harvested Zone – Market Destination (w/ Bell Pepper: Red Variety)

The final component of this assessment is the profitability of this operation with a focus on the red variety of bell pepper. In this case, the general cost pattern remains similar as in the previous case. However, the magnitude of the operational and investment costs is higher, since focus is given towards the implementation of protected technologies. For example, the levels of investments in protected technologies expected in Albuquerque, Raton, Nogales, and Socorro are above the $1 million mark. As alluded to it earlier, investment costs are the driving factor for those areas in which protected technologies are implemented. On the other hand, water cost is the main driver for open-field installations.

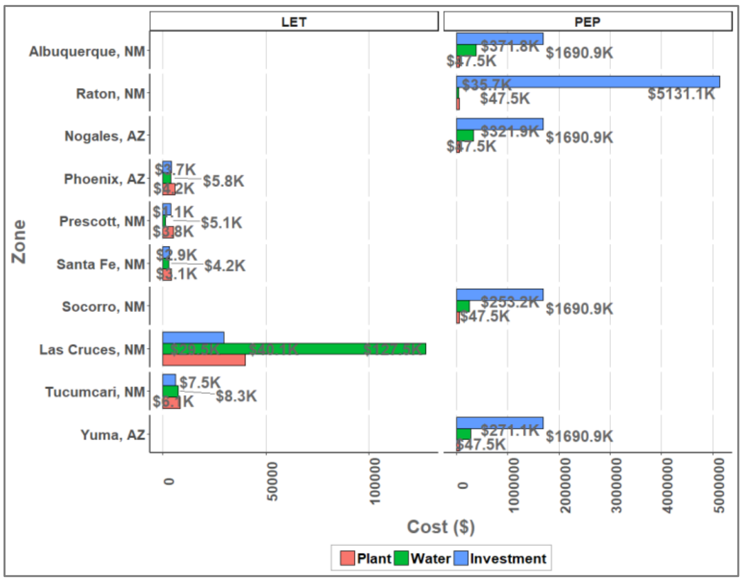


Figure 4-19: Estimated Costs by Type and Zone

Finally, the expected profits and the return on investments for farmers within each region are presented in Table 4-11. From this table, one can observe that there has been a shift toward bell pepper production as expected. The amount of shipments for bell pepper production is higher, attributed to higher revenue streams, while the amount of shipments attributed to lettuce has decreased. Interestingly, the rate of returns for lettuce has remained, while the rate of return for pepper production is relatively higher. In the case of lettuce operations, the reduction in investments can be attributed to a decrease in overall production. For bell pepper production, the shift can be attributed to the focus towards protected technologies that are not as expensive as controlled. Also, since the production does not have to be as high to maintain the level of revenues needed. One should note that although the rate of return for bell pepper investments is higher and the production risk has been lowered through protective technologies, the level of required investment is very high.

Table 4-11: Rate of Return on Investment by Crop (w/Bell Pepper: Red Variety)

Crop	Shipments (1000 lbs.)	Revenues ($1000s)	Costs ($1000s)			Ann ROR
			Production	Logistic	Investment	
LET	235	256	206	2	47	3%
PEP	4678	14943	1491	51	11895	13%
TOM	-	-	-	-	-	-

According to these results, the profitability per participating operation can be improved by concentrating on select product varieties, which can also modify production and harvesting patterns to better match expected market behavior. As it is noted, this is only one of the several options that can be explored by using this framework.

An additional component assessed in the optimization is the impact of problem size on the solving process. To perform this analysis, the number of farmers per regions were varied in order to progressively stress the computational expensiveness of the solution. Table 4-12 the results from these variations on the size of the optimization problem and the types of solving techniques used by the CPLEX solving suite. The first column in this table represents the number of farmers assumed available per region, which serves as the knob for increasing the problem size. In this case, since 11 regions are initially considered in the optimization, each additional farmer per zone actually translates to 11 new entities in the optimization problem. The second and third column represent the number of variables and constraints needed for the formulation, while the fourth and fifth columns represent the number of MIP and branch-and-bound nodes considered while finding solutions. The sixth, seventh, eight, and ninth columns highlight the different types of cuts administered to approximate a final solution. Finally, the tenth and eleventh columns highlight the optimality gap given by the final solution, as well as the total solving time taken by the system.

Table 4-12: Characterization of Optimization Results

Farmers per Zone	Vars	Cons	MIP Iterns	B&B Nodes	Cuts				Gap %	CPU (s)
					Cover	Flow	Gomory	Clique		
1	136,562	41,939	27,899	0	1	2,605	11	1	0.00	4.89
2	147,353	50,134	6,823	0	1	3,905	3	0	0.01	6.44
3	158,144	58,329	9,206	0	1	4,975	2	0	0.01	10.38
4	168,935	66,524	27,311	154	1	5,223	1	0	0.01	23.16
5	179,726	74,719	1323670	5,369	1	4,866	2	0	0.04	298.62
10	233,681	115,694	11379372	3,481	1	5,454	2	0	0.02	511.31
15	287,636	156,669	73388303	37,372	2	6,047	2	0	0.02	4124.73

From this table, one can observe that the optimization formulation is able to handle relatively large problem instances. However, as the number of farmers per region surpass 15 (or 155 total farmers), the solving scheme begins to have issues in finding approximations to the optimal solution. One of the options to reduce the size of the problem could be to reduce the number of potential decision variables. For example, the selection of transportation mode does not change for larger problem sizes. This means that the risk that an even better solution would be obtained under another transportation mode is very low, and therefore we can eliminate set of decision variables to improve its convergence speed. Nevertheless, for purposes of the initial case study of this dissertation, the instance size does not cause major solution issues.

4.7 Discussion

In this section, the development of a practical yield estimation framework is developed, along with a deterministic optimization model, which outputs an optimal combination strategy for the implementation of complementary production systems. There are several benefits to this optimization-based framework. First of all, no optimization tool exists that aims to explore and identify alternative production opportunities for high-value vegetable items based solely on environmental conditions. Specifically, this work provides a supply chain planning perspective for production planning within fresh-food systems by using an analytical framework to incorporate product requirements and temporal market and environmental behavior. Also, this work provides the basic framework for an alternative method of estimating yields of high-value crops based solely on environmental parameters. Furthermore, as mentioned in previous sections, the focus of this work is on the production of high-value vegetables items and the effect of more sophisticated protective technologies (e.g. greenhouse), which differentiates it from existing works in literature that have a much larger scope but focus on less perishable items and less sophisticated technologies.

There are also limitations to the current modeling framework. One of the most important is the refinement of the yield function considered. This includes the possibility of incorporating additional production parameters into this function, such soil properties and amount of sunlight hours received. Again, we stress that the results of the modeling framework would be an initial indication of an opportunity that has been identified but more specific information would be needed to make a finalized investment decision. Another important limitation is the assumption that weekly environmental and price behavior are deterministic, which drives the results to be geared towards the average case. Thus, the incorporation of the stochasticity for these parameters would increase the resolution of the decision-making framework. An additional limitation of this work is the assumption that the local agricultural system would behave as a centralized decision-maker, which in some cases may not be a true representation. Improvement on these limitations may improve our obtained results. In Chapter 5, the stochastic components for this problem is included, while Chapter 6 will be focused on the development of a decentralized formulation that would allow multiple farmers within a given region to share resources and technologies provided by the main investor.

As fresh-food systems continue to grow, it will be important to enhance agricultural planning tools that are geared towards their sustainable and profitable implementation. Through this work, we aim to contribute to this area of research, while also calling on the rest of the community to address supply chain problems specific to micro-farming production. This enables local communities to have a set of decision-making tools for a methodical implementation of local agricultural systems, which in turn results in more efficient operations and improve the chances of success. Applications beyond the study of agricultural research is a potential extension to this work. In the next section, the variability of the stochastic parameters is incorporated into the optimization framework.

5. STOCHASTIC FRAMEWORK FOR COMPLEMENTARY SYSTEMS

A natural extension to the deterministic, formulation developed in the earlier section is to incorporate the stochasticity of environmental and market parameters. The initial step in this extension is to transform the deterministic formulation into a stochastic form such that it exploits the 'wait-and-see' structure of the problem. As mentioned in section 3.5, the local food system problem can be broken into two instances with the same ultimate objective of maximizing profits. The first instance of the problem selects an optimal combination of regions, crops, technologies, and markets to incorporate into a local food system. Once the first instance has been solved, the second instance of the problem is comprised of the set of logistical decisions made once the stochastic parameters have been realized. In this problem, the number of possible parameter realizations can be large, especially when the dimensionality of the set is increased, which makes solving the optimization problem more difficult to solve.

A 'wait-and-see' approach breaks the problem into two parts: the initial first-stage decisions based on a set of known deterministic parameters and the second-stage decisions made once the unknown parameter values have materialized. Within the context of this problem, the set of second-stage parameters are temperature and precipitation values, whose value cannot be predicted exactly in advance and can be represented by probability distribution functions (pdf), say $f(\boldsymbol{\xi})$, where $\boldsymbol{\xi}$ is a multivariate random variable. To find the probability for different values of $\boldsymbol{\xi}$, one could simply calculate $\int f(\boldsymbol{\xi})\, d\boldsymbol{\xi}$, which in turn allows the estimation $E[\boldsymbol{\xi}]$. However, as the dimensionality of $\boldsymbol{\xi}$ increases, the calculation of this integral loses its triviality, especially when the distribution $\boldsymbol{\xi}$ is more complex. Moreover, since the goal is to incorporate these realization as part of an optimization framework, working with integral estimations to find optimal solutions can become burdensome and difficult to solve and sometimes difficult to observe.

To circumvent this problem, a common approach is to discretize $\boldsymbol{\xi}$ into a set of possible second-stage outcomes. By generating many discretized random samples, ω, of the stochastic parameter set, one can reconstruct a representation of its complete sample space, Ω. Within an optimization context, a discretized representation facilitates the use of decomposition-based solving schemes. As it will be discussed in the next section 0, a decomposition approach treats the first and second-stage problems as separate and simpler optimization instances (i.e. master and sub-problem), that individually are easier to solve. Decomposition-based solving schemes iteratively construct supporting hyperplanes to the objective function of the master problem such that no constraint in the overall optimization is violated.

Within a stochastic optimization context, each generated outcome, ω, is treated as a single instance of the second-stage problem. The set of decisions, $\boldsymbol{y}$, optimize an objective function $q(\omega)\boldsymbol{y}$, where $q(\omega)$ is a second-stage parameter vector as a function of the individual event, ω. Second stage decisions, $\boldsymbol{y}$, are assumed to be constrained to the set $S_2 = \{W\boldsymbol{y} = r(\omega) - T(\omega)x, \boldsymbol{y} \geq 0\}$, where $r(\omega)$ and $T(\omega)$ are the parameter vector and matrix, respectively, and dependent on the event ω, x is the current solution from the first-stage problem, and W is a fixed matrix. Note when the event ω is generated, the value of $r(\omega)$, $T(\omega)$, and $q(\omega)$ become known, which are then used to obtain the solution for this instance. Variables in bold represent decision variables whose values have not been solved, while non-bold variables are those that have been solved and have become fixed parameters, such as the case of x in the second-stage sub-problem. Furthermore, as it will be described in the next sections, the solutions from individual subproblems are then used to generate optimality cuts to the master problem. Therefore, by generating multiple scenarios and their respective second-stage solutions, one can approximate the solution of the overall optimization without having to solve the whole problem, as a whole.

5.1 Design for Stochastic Framework

The introduction of variability into an optimization framework for yield exploration has traditionally been approached through the simulation of exogenous weather parameters into a deterministic set-up, similar to the integrated assessment models addressed in section 2.2. Within an agricultural planning framework, Ahumada et al. (2012) introduces variability into the market prices of a stochastic optimization formulation using a deterministic cutting plane algorithm. The focus of this dissertation expands on this approach to assess the problem from an exploratory perspective, in which not only market prices are stochastic, but the expected yields as a function of each region's weather patterns, are also random. This greatly increases the size of the optimization problem, and as noted later in the section, limits its applicability when the number of scenarios is large. Moreover, the use of deterministic cutting plan algorithms may limit the amount of information that can be extracted to characterize the interaction between weather patterns, yields, and investment decisions. To address some of these shortcomings, a stochastic decomposition approach is used to handle larger problem instances. As part of this work, machine learning techniques are used alongside the optimization to learn the interaction between the sampled yield space and technology investment decisions.

The proposed solving scheme design is based on the decomposition of the problem into first and second stage components (Figure 5-1). The first stage formulation is comprised of decisions dependent on parameters whose values are initially known (e.g. the total investment cost under a given technology selection). The second stage components are those decisions made once stochastic parameters (i.e. market price, temperatures, and precipitation) have been observed. The second stage formulation is based on the discretization of the stochastic parameter set, where each random event is solved individually. The feasible space of the overall optimization problem is constructed through an iterative process in which the collection of generated scenarios is used to approximate a solution. Finally, using cutting plane techniques, the dual variables of the individual second-stage problems are used to construct optimality cuts defining the feasible space of the first-stage problem to converge its solution to an optimal value. As noted, another main contribution from this work is the incorporation of a machine learning component that models the interaction between optimization outputs and the characterization of generated scenarios. In the following few sections, each of these components is discussed in the order given by the number in parentheses in Figure 5-1.

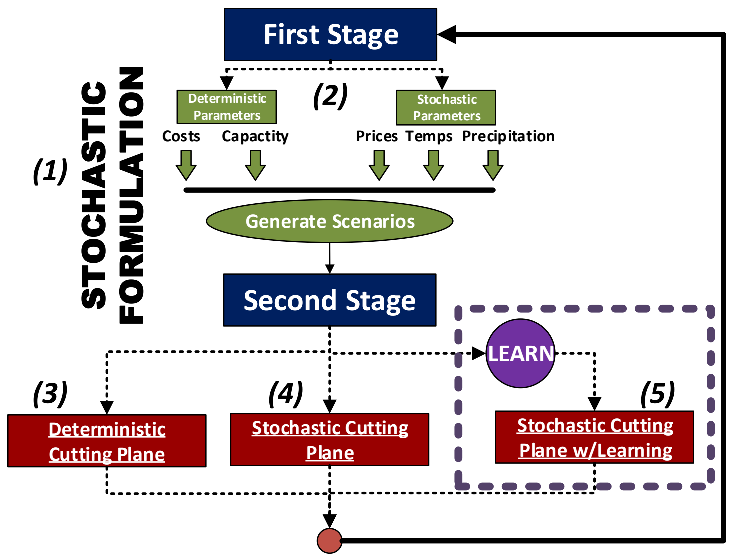


Figure 5-1: Design of Solving Scheme for Stochastic Framework

One of the most important components in the development of the solving scheme design is the process through which cutting planes are constructed and added back into the optimization framework. In addressing this problem, two main cutting plane methods are considered, as shown in Figure 5-1. The first method is based on the construction of deterministic cutting planes, in which a fixed number of discretized scenarios are considered and solved to optimality. The solving scheme used, commonly referred to as the L-Shaped method, is based on a Bender's decomposition approach in which the dual solutions for the complete set of second stage problems (or subproblems) are used to construct first-stage optimality cuts until all first-stage constraints are satisfied. One of the advantages of this approach is that it allows the problem to reach a definite optimal solution for a given scenario subset. However, given that it considers all scenarios from the start, the solving scheme can become difficult to solve when the number of scenarios and the size of the formulations are large. In certain instances, the solution of the solving scheme can be accelerated by allowing each iteration of the algorithm to introduce many optimality cuts, at once, in a process known as the multi-cut L-Shaped method (Birge and Louveaux, 2000). However, a drawback of this approach is that given that the optimization is treated as a single optimization framework, it is difficult to observe the effect from each individual scenario on optimality results.

The second method is based on a stochastic cutting plane approach in which the number of sampled scenarios used by the optimization grows with each iteration. This approach allows the user to observe the interaction between the sampling component of the problem and the optimization framework. This solving scheme approach, known as stochastic decomposition, requires minimal information storage of previously generated instances within the constraints. In turn, it allows the consideration of many more scenarios when compared to the deterministic cutting plane method (Higle and Sen, 1996). However, a potential disadvantage is that its convergence is not dependent on a deterministic value, but rather is based on long-run statistical properties. It also depends on the independency of multiple generated scenarios.

The method proposed in this dissertation is the expansion of the stochastic cutting plane approach by enhancing the connection between the optimization and the sampling components. Using the structure of stochastic decomposition, in which the size of the scenario set grows alongside each iteration, a support vector machine component is inserted into the solving scheme in order to characterize generated scenarios and learn relationships between previously solved instances and the optimization. As part of this process, principal component analysis is used to reduce the dimensionality of the generated scenarios, which facilitates the exploration of these relationships, while also simplifying the process of iteratively training support vector machine models. These models also help assess how well previously generated problem instances represent the full scenario set and estimate whether a new scenario instance will change future optimization values. By combining the reduced dimensionality data set with support vector machines models, one can also synthetically construct yield scenarios that would improve the likelihood of any one investment selection of entering the first-stage solution.

In the following sections, the stochastic formulation is presented as it relates to the context of this problem. Secondly, the mechanism through which discretized scenarios are created is exemplified with analysis examples. Thirdly, the algorithm design for deterministic and stochastic cutting plane mechanisms are shown. APPENDIX A is also available to further detailing the inner workings of implemented techniques. Lastly, the implementation results from these methods are presented along with a comparative analysis. The chapter is concluded with a results discussion and potential areas of future work.

5.2 Stochastic Formulation

The stochastic formulation is broken into two components. The first stage problem is dictated by the set of decisions, $\boldsymbol{x}$, made under deterministic parameters known at the beginning of the 'wait-and-see' strategy. The objective of the first stage problem is to maximize $f(\boldsymbol{x}) = \{c^T\boldsymbol{x} | A\boldsymbol{x} = b, \boldsymbol{x} \geq 0\}$, where matrix A and vectors b and c are known. The second-stage problem is dictated by the set of decisions, $\boldsymbol{y}$, made once the event ω of the random parameter $\boldsymbol{\xi}$ materializes. The objective of the second-stage of the problem is the maximization of the expected value $Q\big(x, \xi(\omega)\big) =$

$\{q(\omega)^T \boldsymbol{y} \mid W\boldsymbol{y} = r(\omega) - T(\omega)x,\ \boldsymbol{y} \geq 0\}$, where $r(\omega)$ and $q(\omega)^T$ are random parameter vectors, $T(\omega)$ is a random matrix that is known once event ω occurs, matrix W is assumed fixed, and x is the solution derived from the first-stage problem. Known as the recourse matrix, having a fixed W assures convexity of the second-stage problem when ξ has finite moments for both linear and non-linear problems (Birge and Louveaux, 2000). The following summarizes the structure of the stochastic formulation:

$$
\begin{aligned}
\text{Max} \quad & z = c^T\boldsymbol{x} + E_\xi\left[Q\left(x, \xi(\omega)\right)\right] \\
\text{s.t.} \quad & A\boldsymbol{x} = b \\
& \boldsymbol{x} \geq 0 \\
\text{where} & \\
Q\left(\boldsymbol{x}, \xi(\omega)\right) = \quad & \max_{\boldsymbol{y}}\{q(\omega)^T\boldsymbol{y} | W\boldsymbol{y} = r(\omega) - T(\omega)\boldsymbol{x}, \boldsymbol{y} \geq 0\}
\end{aligned}
\qquad \textbf{\textit{Eq. 5-1}}
$$

In its continuous form, the estimation of $E_\xi\left[Q\left(x, \xi(\omega)\right)\right]$ requires the integration of a multivariate random variable's, ξ, pdf which can be cumbersome and difficult to obtain. To circumvent this problem, a classical approach is to discretize ξ into a series of k generated scenarios, each with probability p_k of occurring. This allows the estimation of $E_\xi\left[Q\left(x, \xi(\omega)\right)\right]$ through the discretized approximation of $\sum_{k=1}^{K} p_k Q_k(x, \xi(\omega_k)$ where $Q_k\left(\boldsymbol{x}, \xi(\omega_k)\right) = \max_{y_k}\{q_k^T y_k | W y_k = r_k - T_k x, y_k \geq 0\}$. Given that each scenario is independently generated, each optimization of the second-stage problem can be solved separately, which lends itself nicely to the application of decomposition algorithms. The discretized form of the stochastic formulation is summarized in the following:

$$
\begin{aligned}
\text{Max} \quad & z = c^T x + \sum_{k=1}^{K} p_k q_k^T y_k \\
\text{s.t.} \quad & Ax = b \\
& T_k x + W y_k = r_k \quad k = 1, .., K \\
& x \geq 0, y_k \geq 0 \quad k = 1, .., K
\end{aligned}
\qquad \textbf{\textit{Eq. 5-2}}
$$

One should note that the number of generated scenarios, K, remains a problem-dependent parameter. A quick estimation for its value is to use a sample average approximation scheme in which one simply replaces the second-stage objective function in problem *Eq. 5-1* above with a sample average approximation $\frac{1}{N}\sum_{n=1}^{N} q_n^T y_n$, in which statistical convergence properties can be used to obtain an estimate for N (Kleywegt et al., 2002). However, with this approach, the user might run into size issues when attempting to solve optimization requiring large number of samples. This is especially true in a deterministic cutting plane scheme in which the information of every scenario is stored throughout the solving process. Furthermore, the approach may be naïve in some cases. For example, one could potentially characterize the set of scenarios that have been created to gauge the impact that individual scenarios might have on the solution space without having to run the entire scenario set. As discussed later in the chapter, a stochastic decomposition approach, in which sample size grows progressively, allows the user to explore the relationship between the characterization of the observed instances and the results from the optimization framework, which can be used to determine the number of samples.

Following a similar set, variable, and parameter name convention used in section 4.3, the stochastic framework for local food system exploration and development is presented next. As noted in the solving scheme design shown in Figure 5-1, price, temperature, and precipitation are assumed to be random variables. One should note that consequently, yield is also random since its values are directly calculated from temperature values as explained in section 4.2. In the stochastic formulation, the expected value for the second-stage problem has been discretized into K scenarios and considered independent subproblems.

The solution of the master problem (i.e. first-stage problem), $x_{jfu}^{t_p}$, is used to solve each second-stage subproblem. Furthermore, as it will be noted, constraint $\boldsymbol{n}(\boldsymbol{x}) \geq \boldsymbol{\alpha_s} + \boldsymbol{\beta_s x}$ in the master problem is used as a placeholder for optimality cuts constructed via the different solving schemes.

Next, the formulations for the master problem and a single instance, k, of the subproblem are described:

Sets:

$z \in Z$:	Set of zones (regions)for production
$f \in F(z)$:	Set of farmers in zone z
$d \in D$:	Set of distribution centers
$j \in J$:	Set of crops
$c \in C$:	Set of customers in different markets
$m \in M$:	Set of transportation modes
$t \in T$:	Set of time planning periods
$t_p \in T_p \subset T$:	Set of planting periods in T
$t_h \in T_h \subset T$:	Set of harvesting periods in T

Decision Variables:

B_{jfu}:	$\begin{cases} 1 & \text{if tech } u \text{ is available to farmer } f \in F(z) \text{ for crop } j \\ 0 & \text{otherwise} \end{cases}$
$X_{jfu}^{t_p}$:	Yld of crop j by farmer f when planted at t_p using technology u
$MicroHarv_{jz}^{t_h}$:	Amount of crop j harvested during t_h within zone z
$Pack_{jz}^{t_p,t_h}$:	Amount of crop j packaged during t_h planted in t_p within zone z
$WA_z^{t_p,t_h}$:	Additional water allocated to region z between t_p and t_h *(if rainfaill is not enough)*
$SLZ_{jfz}^{t_h,t}$:	Qty. shipped of crop j from farmer m *to* region z at time t harvested at t_h
$SZD_{jzdm}^{t_h,t}$:	Qty. shipped of crop j from region z *to* DC d at time t harvested at t_h on mode m
$SDC_{jdcm}^{t_h,t}$:	Qty. shipped of crop j from DC d *to* market c at time t harvested at t_h on mode m
$\mathrm{Invw}_{jz}^{t_h}$:	Inventory of crop j at zone z at time t_h

$\mathrm{Invw}_{\mathrm{jz}}^{\mathrm{t_h}}$: Inventory of crop j at zone z at time $\mathrm{t_h}$
$AddWCap_z$ Additional warehouse capacity used at zone z

Parameters:

$Land_f$: Land available to farmer f
$yield_{jzu}^{t_p,K}$: Yld of crop j in zone z planted in t_p using tech u for problem K
$YDis_{jzu}^{t_p,t_h}$: Yld distribution planted/harvested in t_p/t_h using tech u for crop j in region z
$lrainrec_z^{t_p,t_h,K}$: Rain received between t_p and t_h in region z for problem K
$WReq_{ju}$: Water requirements for crop j using technology u
$MaxDem_{jm}^t$: Maximum demand for crop j by customer m at time t
$MinDem_{jm}^t$: Minimum demand for crop j by customer m at time t
$maxl_j$: Maximum number of acres that can be planted of crop j
$minl_j$: Minimum number of acres that can be planted of crop j
Mpr_{jm}^t: Price offered for crop j by customer m at time t
$Ctech_u$: Amortized investment cost of technology u
$Cplant_j$: Cost of planting a full acre of crop j
$Coper_u$: Cost of operating technology u for one year
$CTZC_{zcm}$: Transportation cost from region z to market c using mode m
$CTDC_{dcm}$: Transportation cost from DC d to market c using mode m
$CTZD_{zd}$: Transportation cost from zone z to DC d
$CTLZ_z$: Transportation cost within zone z
Cw_z: Inventory cost at zone z
Cd_d: Inventory cost at DC d
$Ccase_j$: Packaging cost for crop j
$Cwater_z$: Cost of water in region z
$Cavail$: Capital available to investor for investment in all regions
$Time_{zm}$: Transportation time between zone z to market m

Master Problem:

The master problem is comprised of the set of decisions made during the first-stage problem. This set of decisions consist of mainly the number of acres to plant within each region by farmers involved and includes a selection of technologies to implement within each region. Technology costs are divided into an annualized investment cost, $Ctech_{uz}$, and annual operational cost $Coper_{uz}$ in the same manner as they were defined in the deterministic version. As part of this formulation, we have included the estimate of the recourse function as $\boldsymbol{\eta}(\boldsymbol{x})$, whose definition varies according to the solving method used and is approximated via the optimality cuts defined by constraint Eq. 5-3 below. Hence, the goal of the solving methods is a convergent approximation of the estimated second-stage value.

Maximize:

$$- \sum_{t_p jfu} X_{jfu}^{t_p} * CPlant_j - \sum_{t_p jfuz: f \in F(z)} X_{jfu}^{t_p} * [Ctech_{uz} + Coper_{uz}] - \boldsymbol{\eta}(\boldsymbol{x}) \qquad \textbf{Eq. 5-3}$$

Subject to:

$$\sum_{juf: f \in F(z)} B_{jfu} * Ctech_{uz} \leq Cavail \qquad \textbf{Eq. 5-4}$$

$$\sum_{t_p} X_{jfu}^{t_p} \leq Land_f * B_{jfu} \quad \forall f, j \in J, u \in U \qquad \textbf{Eq. 5-5}$$

$$\sum_{uf} B_{juf} \leq CropOper_j \quad \forall j \qquad \textbf{Eq. 5-6}$$

$$X_{jfu}^{t_p} \leq maxl_j * B_{jfu} \quad \forall t_p \in T_p, j, f, u \qquad \textbf{Eq. 5-7}$$

$$\sum_{t_p u} X_{jfu}^{t_p} \geq minl_j * \sum_{u} B_{jfu} \quad \forall t_p, j, f \qquad \textbf{Eq. 5-8}$$

$$\boldsymbol{n}(\boldsymbol{x}) \geq \boldsymbol{\alpha}_s + \boldsymbol{\beta}_s \boldsymbol{x} \quad \forall s \in CUTS \qquad \textbf{Eq. 5-9}$$

Besides constraint Eq. 5-9, the other objective and constraint formulations do not change from the deterministic version described in section 4.3. As mentioned, constraint Eq. 5-9 is comprised by the set of optimality cuts defined by $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, which is constructed using sub-problem dual solutions and first-stage solution variable $\boldsymbol{x}$. Note that their actual construction will be part of the three cutting plane algorithms described in sections 5.4, 5.5, and 5.6. After solving the master problem, the resulting value of $\boldsymbol{x}$ becomes a fixed parameter to be used by each of the subproblems. Next, the construction of an individual subproblem is shown. Note that the integration of the results of each subproblem into the overall optimization is dependent on the cutting plane algorithm used and explained subsequent sections.

Sub-Problem:

The formulation of the sub-problem is based on the set of decisions one would make under a single, discretized instance of the second-stage where each k sub-problem refers to a generated scenario of the random parameters, $lrainrec_{z}^{t_p,t_h,k}$, $yield_{jzu}^{t_p,k}$, and $Mpr_{jc}^{t,k}$. In this formulation, $x_{jfu}^{t_p}$ refers to the optimal solution from the first-stage, or master, problem, used as input for the second-stage problem. Furthermore, to assure feasibility, the warehouse capacity constraint Eq. 5-15 is modified to assure feasibility by heavily penalizing the objective function whenever there is an unfeasible requirement. In later sections, an explanation is given on how each cutting plane algorithm uses the sub-problem dual solutions to generate the optimality cut $\boldsymbol{n}(\boldsymbol{x}) \geq \boldsymbol{\alpha} + \boldsymbol{\beta x}$ within the master problem. The formulation for sub-problem k is presented next.

Maximize $(\forall \boldsymbol{k} \in \boldsymbol{K})$***:***

$$\sum_{jdmt_htc:t=t_h+LT_{dc}} SDC_{jdcm}^{t_h,t} * Mpr_{jc}^{t,k} - \sum_{jqzt_ht} Invw_{jz}^{t_ht} * Cw_z - \sum_{jqdt_ht} Invd_{jd}^{t_ht} * Cd_d - \sum_{t_pt_hz} WA_{z}^{t_pt_h} * Cwater_z - \sum_{z,t_h,j} Pack_{z,j}^{t_h} * Ccase_j \qquad \textbf{Eq. 5-10}$$

$$- \sum_{jdmt_htc} SDC_{jdcm}^{t_h,t} * CTDC_{dcm} - \sum_{jzt_htd} SZD_{jzdm}^{t_h,t} * CTZD_{zd}$$
$$- \sum_{jft_htz:f\in F(z)} SLZ_{jfz}^{t_h,t} * CTLZ_z$$
$$-M * \sum_{z} AddWCap_z$$

Subject to:

$$WA_z^{t_p,t_h} \geq -lrainrec_z^{t_p,t_h,k} + \sum_{jfu:f\in Z(f)} WReq_{ju} * x_{jfu}^{t_p} \quad \forall t_p, t_h \in T_h \; z \in Z \qquad \textbf{Eq. 5-11}$$

$$MicroHarv_{jz}^{t_h} = \sum_{t_p} x_{jfu}^{t_p} * YDist_{jzu}^{t_p,t_h} * yield_{jzu}^{t_p,k} \quad \forall t_h, j, z, u, \; f \in F(z) \qquad \textbf{Eq. 5-12}$$

$$SLZ_{jfz}^{t_h,t} = \sum_{u:q=2,f\in F(z)} MicroHarv_{jfu}^{t_h} \quad \forall t_h, j, z, \; f \in F(z) \qquad \textbf{Eq. 5-13}$$

$$Pack_{jz}^{t_h} = \sum_{fu:q=2,f\in F(z)} MicroHarv_{jfu}^{t_h} / ContCap_j \quad \forall t_h, j, z \qquad \textbf{Eq. 5-14}$$

$$\sum_{jq:q=2} Pack_{jz}^{t_h} \leq WZCap_z + AddWCap_z \quad \forall z \qquad \textbf{Eq. 5-15}$$

$$Invw_{jz}^{t_ht_h} = \sum_{f\in F(z)} SLZ_{jfz}^{t_ht_h} \quad \forall t_h, j, z \qquad \textbf{Eq. 5-16}$$

$$Invw_{jz}^{t_ht} = Invw_{jz}^{t_ht-1} + \sum_{f\in F(z)} SLZ_{jfz}^{t_ht} - \sum_{d} SZD_{jzdm}^{t_h,t} \quad \forall t_h, t, j, \; t > t_h \qquad \textbf{Eq. 5-17}$$

$$Invd_{jd}^{t_ht_h} = \sum_{z} SZD_{jzdm}^{t_h,t} - \sum_{cm} SDC_{jdcm}^{t_h,t} \quad \forall t_h, j, d \qquad \textbf{Eq. 5-18}$$

$$Invd_{jd}^{t_ht} = Invd_{jd}^{t_ht-1} + \sum_{z} SZD_{jzdm}^{t_h,t} - \sum_{cm} SDC_{jdcm}^{t_h,t_h+LT_{dc}} \quad \forall t_h, t, j, d, \; t > t_h \qquad \textbf{Eq. 5-19}$$

$$\sum_{t_hzqmc:t_h+SL_j\geq t\geq t_h} SDC_{jdcm}^{t_h,t_h+LT_{zc}} \leq MaxDem_{jc}^{t} \quad \forall jct \qquad \textbf{Eq. 5-20}$$

$$\sum_{t_hzqmc:t_h+SL_j\geq t\geq t_h} SDC_{jdcm}^{t_h,t_h+LT_{zc}} \geq MinDem_{jc}^{t} \quad \forall jct \qquad \textbf{Eq. 5-21}$$

As with the formulation of the master problem, the constraint formulations for a single, sub-problem instance remain the same as in section 4.3. The only difference is the added

k index in the stochastic parameters. Following the order numbering in Figure 5-1, the next section details how the scenarios for each of the stochastic parameters were constructed. Within the context of the stochastic formulation, each generated scenario constitutes an individual sub-problem.

5.3 Scenario Generator for Stochastic Parameters

Incorporating a scenario generating component to the stochastic framework is an important piece of the overall design. The goal of the scenario generating component is to create a mechanism through which the variability of the stochastic parameters is represented within our stochastic formulation. As mentioned earlier, there are three sources of variability to the problem, temperature, precipitation, and market price. The first source of variability directly leads to the stochasticity of crop yield patterns within each geographical zone and will be discussed at length during this section. The second stochastic parameter, precipitation, dictates the availability of water within each zone, and although it is assumed that farmers will be able to satisfy crops' water requirements, second-stage decisions are still constrained by its availability through the amount of precipitation received. For example, if a region receives sufficient precipitation throughout its planting and harvesting period, then it reduces water costs incurred from sourcing additional amounts. On the other hand, if a zone observes low precipitation values, then the costs of satisfying crop water requirements will rise, which in turn alters the final solution. Finally, the third source of variability is market price, which directly affects the objective function of second-stage subproblems. By combining all components into a single framework, the optimization outputs an optimal strategy seeks to protect production through variable reducing technologies, as well as potentially taking advantage of profit opportunities provided by volatility in market prices.

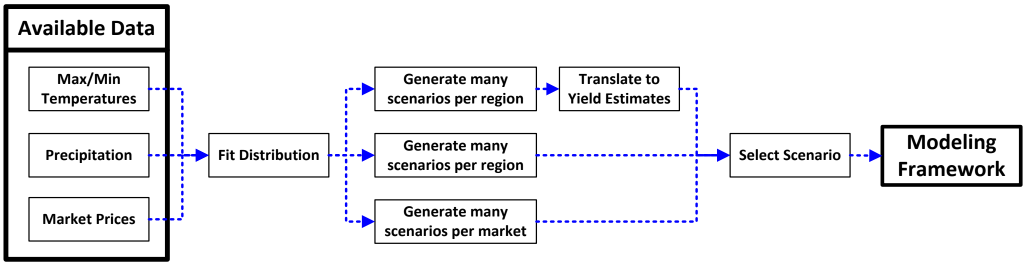


Figure 5-2: Schematic for Generating Scenarios

The schematic through which the scenarios of each stochastic parameter are generated is described in Figure 5-2**.** Although generating scenarios for each stochastic parameter requires a different approach, the general schematic remains the same. The first step is to collect available historical data for each component and used to approximate its distribution. For example, maximum and minimum temperatures, as well as precipitation values, could be obtained from the U.S. National Climatic Data Center over a 30-year period (NOAA 2016a), while average daily market price information can be obtained from USDA (2016) starting from its earliest available year in 1998. The second step in this overall schematic is to fit a probability distribution to each of the random parameters with the purpose of characterizing its behavior through a single multivariate variable and capable of representing weekly behavior for the whole year. This allows the user to generate enough yearly scenarios for each parameter. The third step in the schematic is to translate the temperature values to crop yield estimates such that they can be incorporated into the optimization framework, while precipitation and market price values can be introduced directly. Finally, throughout this section, example analysis and assumption tests are provided to show the steps taken in constructing scenarios.

5.3.1 Generating Yield Scenarios

For purposes of this framework, the main environmental parameters considered in generating yield scenarios are maximum and minimum temperatures based on the estimation methodology presented in section 4.2. Based on empirical data collected from the different regions, it is initially hypothesized that weekly maximum and minimum temperatures throughout the year can be represented as a multivariate normal random variable. This will allow us to represent yearly temperature values as a single multivariate variable that can capture autocorrelation among its weekly values. Also, if one assumes generated temperature scenarios for each region to be independent from one another, then temperature scenarios per zone can be represented through a correlated multivariate variable comprised of 104 weekly maximum and minimum values $w_{ij}^{z} \sim N\left(\mu_{ij}, \Sigma\right): i \in 1.,,52, j \in \{max, min\}$. The multivariate random vector would take the form:

$$\boldsymbol{W} = \left[w_{1,max}, w_{1,min}, \dots,, w_{52,\mathrm{max}}, w_{52,min}\right] \qquad \textbf{Eq. 5-22}$$

To create a multivariate normal fit representative of weekly maximum and minimum temperatures, daily temperature values were used instead of summarized weekly averages. This allows us to increase the number of available observations per week by placing historical daily temperature observations into separate weekly bins. Figure 5-3 shows a schematic of how the different days of the year are split into different weekly bins. Daily maximum and minimum temperature values within each week are then used to estimate μ_{ij} and Σ.

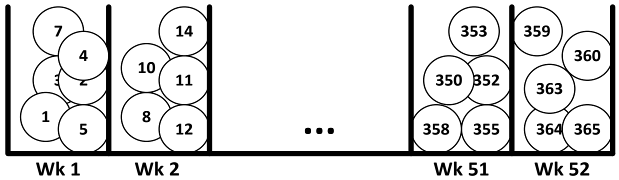


Figure 5-3: Daily Temperatures into Weekly Bins

The equivalent data structure is captured by Table 5-1, in which daily data points per year are sectioned by week.

Table 5-1: Temperature Data Structure for a Single Region

	Week 1		Week 2		…	Week 52	
	Max T	Min T	Max T	Min T		Max T	Min T
Year 1	$d^1_{1,max}$	$d^1_{1,min}$	$d^1_{14,max}$	$d^1_{14,max}$		$w^1_{365,max}$	$d^1_{365,max}$
…							
Year N	$d^N_{7,max}$	$d^N_{7,min}$	$d^N_{14,max}$	$d^N_{14,max}$		$d^N_{365,max}$	$d^N_{365,max}$
	$\bar{\mu}_{1,max}$	$\bar{\mu}_{1,min}$	$\bar{\mu}_{2,max}$	$\bar{\mu}_{2,min}$		$\bar{\mu}_{52,max}$	$\bar{\mu}_{52,min}$

To test the multivariate normality assumption within each zone, the individual weekly assumption was tested first by comparing empirical and theoretical distribution fits and assessing constructed Q-Q plots. Figure 5-4 provides an example snapshot of the normal distribution fits taken for maximum and minimum temperatures of WEEK1 of the planting period in an arbitrarily selected zone (Santa Fe, NM). It is important to note that the single normal distribution fits for other weeks and zones exhibit similar empirical and theoretical characteristics. From these plots, one can deduce that the normality assumptions are satisfactory even though there are some minor indications of positive skewness of the empirical distribution for minimum temperatures as shown by the downward tendency of Q-Q plot fits.

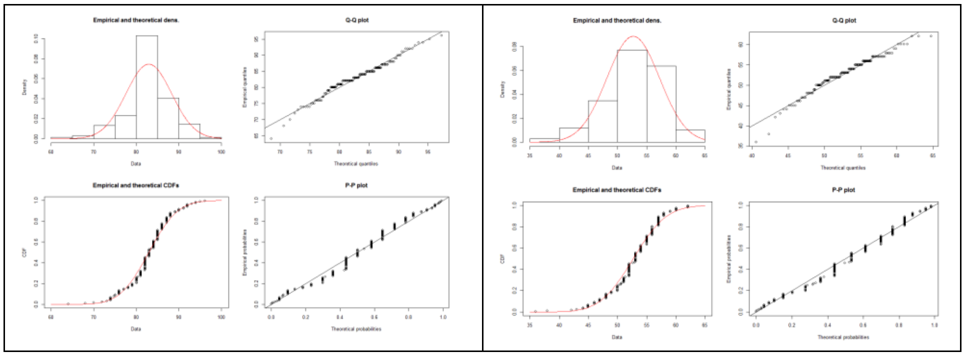


Figure 5-4: Single Normal Distribution Fits for Week 1 (Santa Fe, NM)

To test the multivariate normality hypothesis, Q-Q plots were used. Figure 5-5 provides a sample snapshot of the Q-Q plots of yearly weekly temperatures in Las Cruces, NM, and Yuma, AZ. Again, this general behavior is exhibited within the Q-Q plots of the rest of the zones. Based on the Q-Q plots, one can deduce that the multivariate assumption for temperatures is satisfactory given the dimensionality of the multivariate variables. Nonetheless, the deviating tails on the Q-Q plots signify heavy tails on the normality assumption, which may signify some violation of the multivariate normality assumption, as well, as potential issues with the available dataset. However, although these fits do not perfectly align with the theoretical distribution, it does allow a satisfactory representation of yearly behaviors through a single multivariate normal random variable, which meets one of the goals of the scenario generating component of this stochastic framework. If the normality assumption is adequate, another option is the use non-parametric distributions to fit a probability density function on the empirical information. However, for the purposes of our case, multivariate normality assumption is satisfactory given the large number of variables.

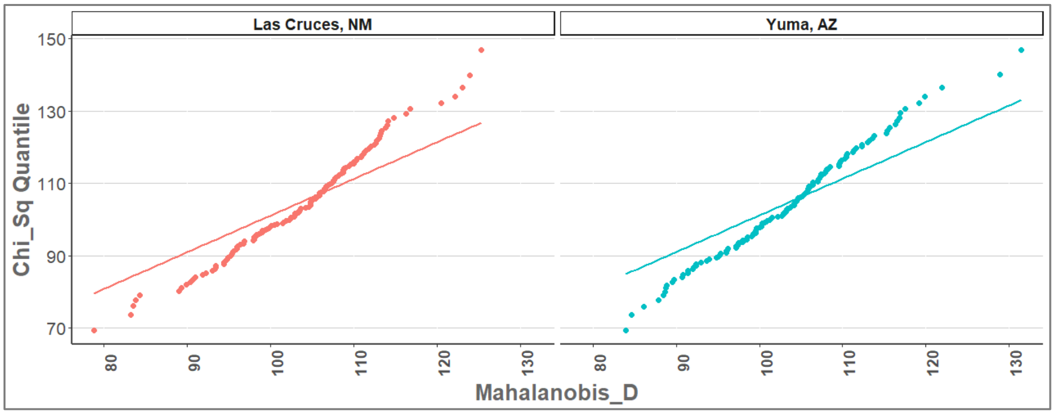


Figure 5-5: Q-Q Plots for Multivariate Normal Distribution Fits

The last component in generating temperature scenarios is the assumption that weekly maximum and minimum temperatures are correlated. To generate correlated random variables, one can use Law and Kelton (1991) methodology in which a Cholesky decomposition of the covariance matrix **Σ** is used. Then using the Cholesky decomposition matrix, one can generate correlated values through the following procedure:

$$w_{i,j} = \bar{\mu}_{i,j} + \sum_{k=1}^{i} c_{kj} * \bar{\mu}_{j,max} \; ; \forall \; i \in \{1,..,52\}, j \in \{min, \max\}; \qquad \textbf{Eq. 5-23}$$

$$where \; c_{kj} \; is \; the \; (k, j) \; component \; of \; the \; Cholesky \; matr$$

This method allows one to generate an arbitrary number of temperature scenarios, where each scenario is comprised of weekly maximum and minimum temperature values for one year. For example, Figure 5-6 graphs maximum and minimum temperatures for years 2014 through 2016 within Santa Fe, Tucumcari, and Yuma. In this figure, maximum temperatures are represented by solid lines, and minimum temperatures are labeled in dashed lines. Each color represents a different year, while the black lines represent the two generated scenarios. By comparing the actual temperatures versus the generated ones, one can observe that the scenarios follow the general behavior of the observed temperatures.

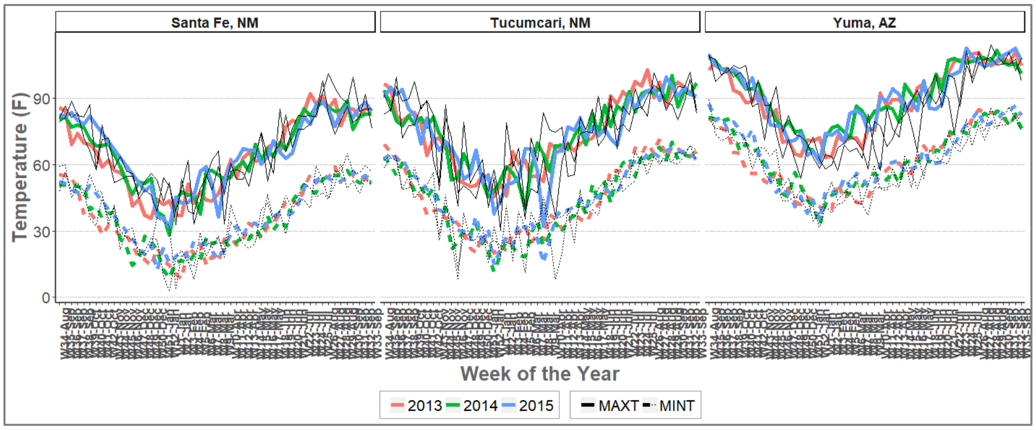


Figure 5-6: Generated Max and Min Weekly Temperatures per Year

The final step into the scenario generating process is to transform these temperature vectors into actual yield estimates, which would then be inputted into the optimization. In the deterministic representation discussed in section 4.3, the average weekly temperature vector was used to estimate crop yields under varying temperature scenarios. In the case of unprotected production, yield estimates depend completely on the observed temperature. However, under protective technologies, temperatures can be altered, which in turn modifies the expected yield pattern. Under completely controlled environments, one assumes that the temperatures can be kept at the exact temperature needed to maximize yields (Mattson 2015). Therefore, depending on different scenarios of weekly temperatures throughout the year, one can delineate different yield projections as shown next in Figure 5-7.

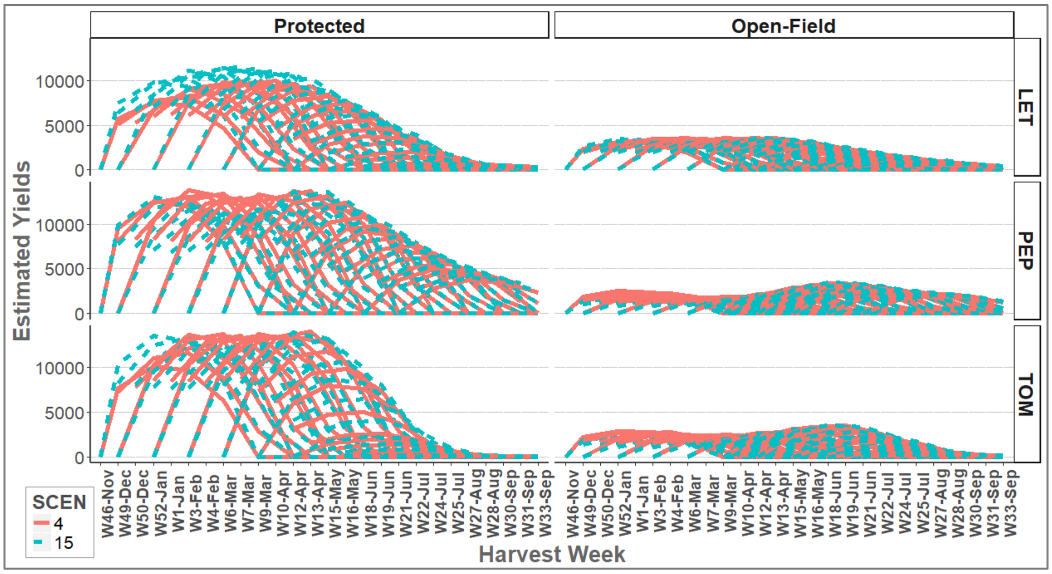


Figure 5-7: Crop Yields Under Two Scenarios in Yuma, AZ

Figure 5-7 shows two scenarios for weekly temperature vectors when translated to yields of romaine lettuce, bell peppers, and tomatoes in Yuma, AZ. On the left-hand side of this figure, one can observe the calculated yields under protected technologies, while on the right-hand side one can observe those for open field. On each level of this figure (from top-to-bottom) is the harvest amount for each crop during different weeks of the year; each line represents a different plant week. For example, the first red line from left to right within each figure is the harvest yield for that crop when planted in WEEK 1. Therefore, from this graph, one can observe how the shape of the harvested yield varies across time assuming different plant weeks. Also, within this graph, one can observe the dashed blue lines, represent another generated scenario of weekly temperatures. By comparing the dashed blue and solid red lines, one can observe how the different weekly temperatures also influence weekly harvested yield. Lastly, one of the things to note in this figure is the yield comparisons between protected technologies and open-field implementations. One can observe the assumption that the expected yields under protected technologies are higher when temperature is kept form reaching outliers and plants are more closely taken care of.

5.3.2 Generating Precipitation Scenarios

Generating precipitation scenarios is different than for weekly temperatures since variables are not normally distributed. This is in part due to the fact that there are restrictions on its values, such as non-negativity constraints and the intermittence of events (i.e. precipitation is not observed every week and cannot be negative). In this case, a combination of binomial and gamma distribution fits are used to generate weekly scenarios with the same strategy as for temperatures. First, daily precipitation values for each region are split into weekly bins following the schematic shown in Figure 5-3 and data structure depicted in Table 5-1. However, instead of using precipitation values directly, we first want to estimate the expected number of days per week were rain would be observed, as shown in Figure 5-8. In this figure, the dark colored circles represent the days for each week that precipitation would be observed. This method allows us to approximate the number of filled circles per weekly bin by using a binomial distribution fit.

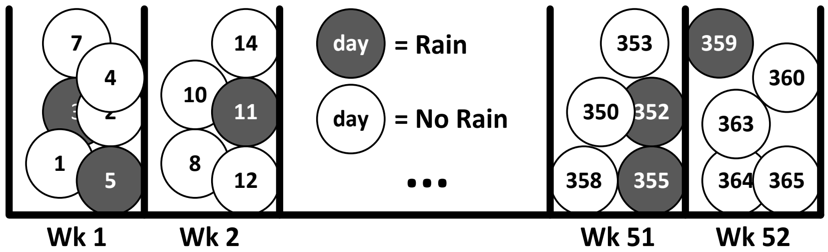


Figure 5-8: Daily Precipitation into Weekly Bins

First, the daily precipitation values per weekly bin are transformed into a binary indicator on whether any day of the week received precipitation. Then, a binomial distribution fit is tested on the transformed data set. Figure 5-9 depicts the empirical versus theoretical binomial distribution fits for WEEK 1 of the planting period in Las Cruces, NM. In this case, one can observe that the theoretical distribution closely follows the behavior of the empirical behavior. The important note in this case it that it allows the generalization of the number of days that observe precipitation within each week. In this manner, one can generalize the precipitation behavior for each week, as well as estimate the probability parameter for this fit.

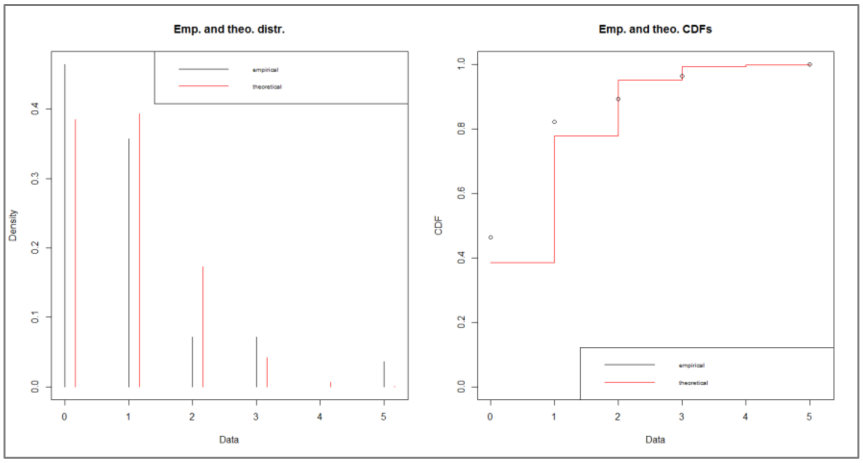


Figure 5-9: Theoretical versus Empirical in Week 1 (Las Cruces, NM)

The final component in this set of fits is being able to estimate the amount of precipitation received given that rain was observed in a given week. A gamma distribution was fitted and tested on the data points whenever rain was received. Figure 5-10 presents a comparison between the empirical and theoretical fits for precipitation values during the first week of the planting period Las Cruces, NM. As one can observe, the empirical distribution of the data closely follows the theoretical values given by the theoretical values of the gamma distribution.

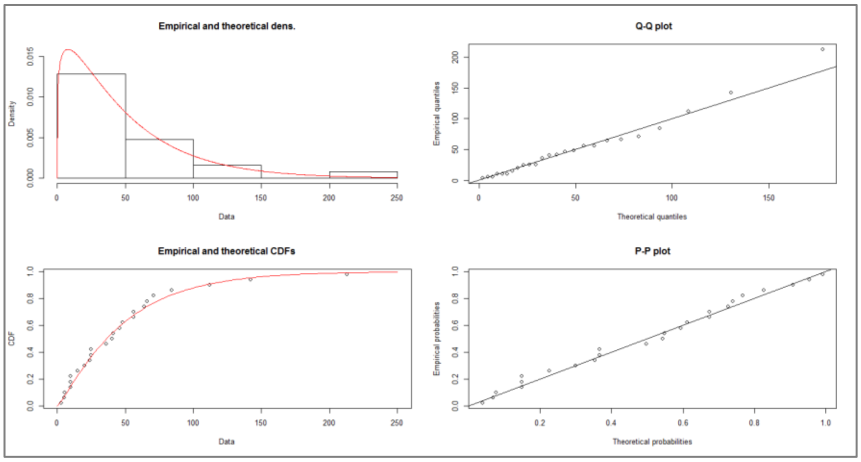


Figure 5-10: Gamma Fit on Prec. Values in Week 1 (Las Cruces, NM)

To generate scenarios for a week, we first generate a scenario for the number of times that precipitation was observed. To estimate the amount of water received via precipitation, we then use the estimate for precipitation (given that rain was observed) using a random variable generator with a gamma distribution. For example, to generate a scenario of precipitation for a given week, one can use:

$$E_{wk_i}[PRCP] = E_{wk_1}[PRCP|RainRec] * P(RainRec) \quad \textbf{Eq. 5-24}$$

Using this approach, one can generate precipitation values for all weeks within a given scenario. Note that in this case, it is difficult to develop a multivariate random variable that would still hold the characteristics of the empirical distribution. Moreover, one would still have to generate random scenarios.

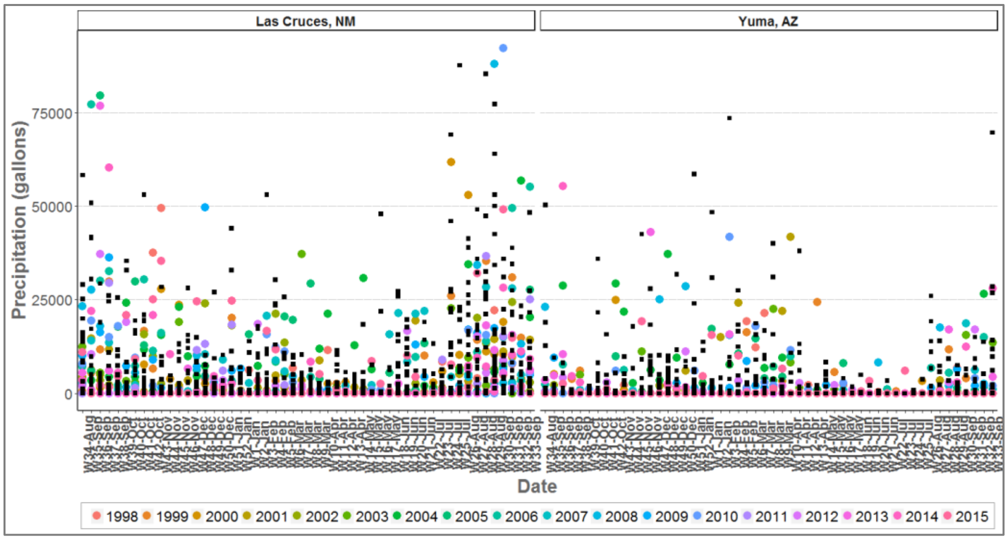


Figure 5-11: Twenty Precipitation Scenarios versus Actual

Figure 5-11 shows twenty generated precipitation scenarios in Las Cruces, NM, and Yuma, AZ, through this simple approach. These scenarios are marked in blacked squares. This figure also contains the actual overserved precipitation values through different weeks of the year. As one can observe from this figure the generated precipitation scenarios can capture the general behavior of the actual system. This allows one to generalize the behavior of yearly precipitation values by generating an arbitrary number of scenarios. The goal is to eventually determine a 'good' number of scenarios that can capture the stochastic behavior of precipitation values.

5.3.3 Generating Market Price Scenarios

The final stochastic component of the framework are the market prices for the different crops within each market. To generate scenarios for this component, a similar approach was used as for temperature values. However, in this case, it was assumed that average market prices were distributed lognormally, since they are nonnegative values. To test this, weekly market price values were assigned to different weekly bins using the schematic given by Figure 5-3 and similar in data structure as Table 5-1. Then a lognormal distribution was fitted and assessed for each week, market, and crop.

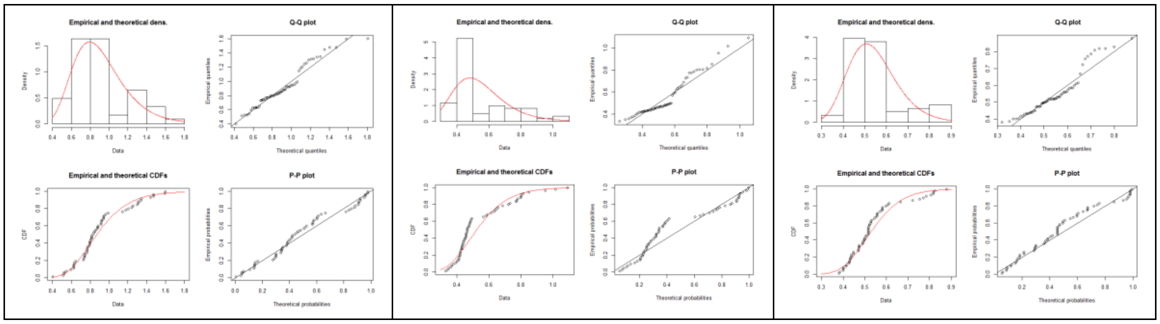


Figure 5-12: Lognormal Dist. for Week 1 in Boston, MA (LET/PEP/TOM)

Figure 5-12 shows the lognormal distribution fits for week 1 of the planting period and for the three crops in consideration. As one can observe, the lognormal distribution fit is not as clean as it was for normal distribution fits for regional maximum and minimum temperatures. For example, the lognormal distribution fits *do* seem to have some issues regarding its kurtosis. This may be due to several reasons including the accuracy of the price data itself given that this information is reported and is open to inaccuracies. Nonetheless, in general the theoretical distribution appears to follow the observed distribution of the data points relatively well when compared to other theoretical distribution types.

Similar to the temperature vector, the aim is to represent weekly prices for the whole year as a single multivariate random vector. Therefore, data values were converted to a normally distributed equivalent and represented as multivariate normal distribution. To test this assumption, a multivariate normal distribution was fitted on the empirical data. In Figure 5-13, one can observe the Q-Q plot for multivariate normal fit on the transformed data. From this figure, one can observe serious deviations from the straight line, which indicates violations for the normality assumptions. However, although the Q-Q plot still show violations from the normality assumption, the plots still show a certain level of satisfactory skewness, although it kurtosis is large. This indicates fairly heavy tails on the normal distribution fits along a fairly symmetrical distribution. Nonetheless, again, this allows us to have single generated values that would represent the weekly values of the whole year with some degree of multivariate normality violations.

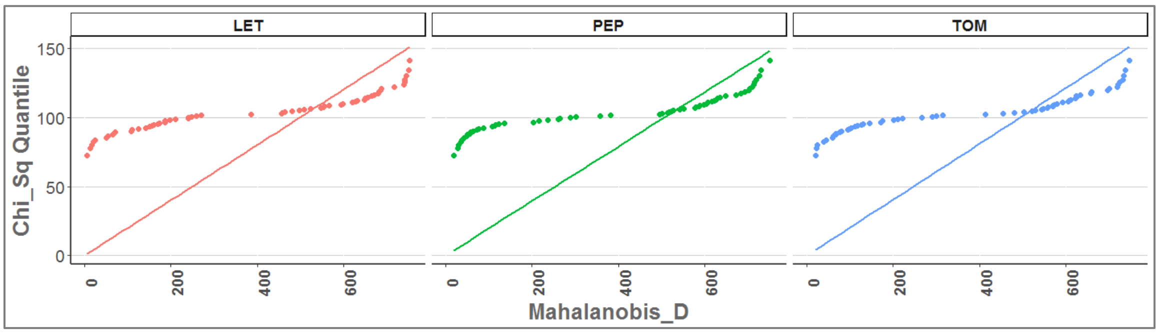


Figure 5-13: Q-Q Plots for Multivariate Normal Fit on Prices (Boston, MA)

Using the same Cholesky decomposition approach as in section 5.3.1, a correlated multivariate normal random variable was created. Then after normally distributed values were generated, these were transformed to their lognormal equivalent, which can generate many scenarios for the different crops and markets.

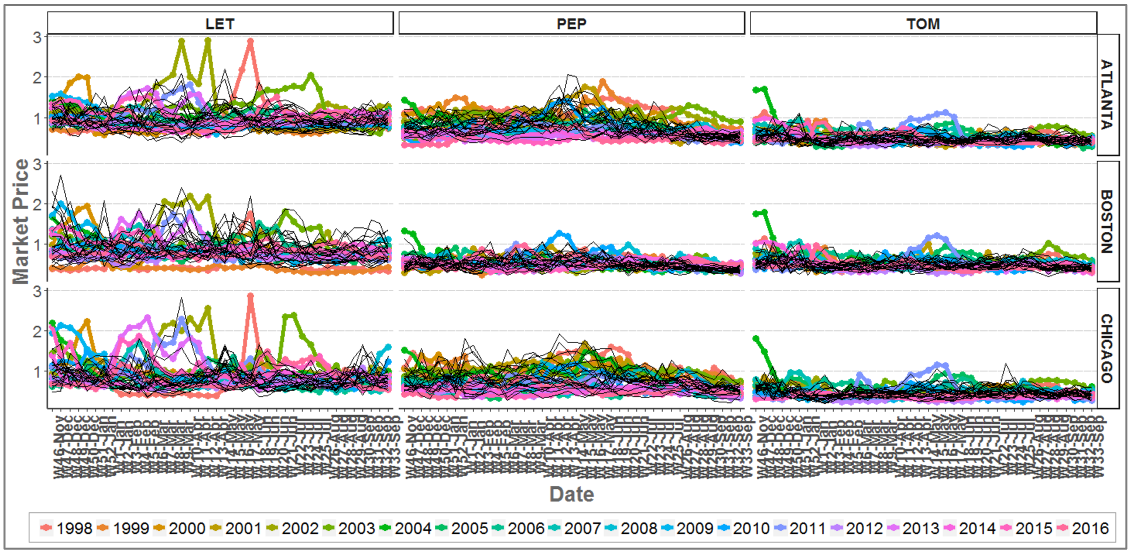


Figure 5-14: Twenty Market Price Scenarios versus Actual

Figure 5-14 presents twenty of these scenarios for lettuce, pepper, and tomato within the markets of Atlanta, Boston, and Chicago. In this figure, the colored lines represent a different year of weekly market prices, while the black thin lines represent a different generated random vector of prices. From this figure, one can observe that the black lines can capture most of the variability observed within the different markets. However, there are certain markets that would need additional generated scenarios to be able to capture the volatility observed during the last twenty years, such as in the case of lettuce in the Atlanta and Chicago markets. The idea is to capture the volatility of market prices through the construction of these scenarios.

After developing the methodology on how to generate yield, precipitation, and market price scenarios, the next step in the meta-design of Figure 5-1 is using this information within the optimization framework. In the next few sections, the solving schemes, deterministic, stochastic, and stochastic while learning cutting planes, are explained in more detail. This includes detailing the specific solving scheme used within each specific approach. Then, a comparison is made between the solving schemes by using their implementation results. Most importantly, we address how the deterministic solution is affected by the incorporation of parameter stochasticity.

5.4 Solving Scheme using Deterministic Cutting Planes

A technique commonly used to address the decomposable, discretized structure shown in problem Eq. *5-2* in section 0 is commonly referred to as the L-shaped method. This algorithm provides a deterministic solution by using a Bender's decomposition approach that solves a series of discretized sub-problems by iteratively generating deterministic cutting planes to approximate the recourse function $\boldsymbol{\eta}(\boldsymbol{x})$ appearing in the objective of Eqn. Eq. 5-3. One of the advantages of such an approach is that it is highly robust to multiple types of convex and nondifferentiable optimization problems (Birge and Louveaux, 2000). However, when the dimensionality of the random parameters increases; the estimated number of scenarios may become large (Kleywegt et al., 2002). Furthermore, since all of the subproblem information is kept throughout the optimization algorithm, it leads to memory constraints of the system when the number of scenarios is large (Higle and Sen, 1996).

The main solving structure of an application of a standard L-Shaped method is summarized by Figure 5-15. The first step is to solve the master problem. Based on the results from the master problem, each instance of a predefined set of scenarios is solved individually. Within the context of the stochastic framework, the formulation used for the master and sub-problems are those detailed in section 0. The second step is to construct a single optimality cut by using the dual solutions generated by every sub-problem. Once the optimality cut has been produced, one assesses if the optimality requirements have been met. Note that in this framework feasibility constraints are not created. Instead, large penalization values are given to infeasible solutions to avoid their selection.

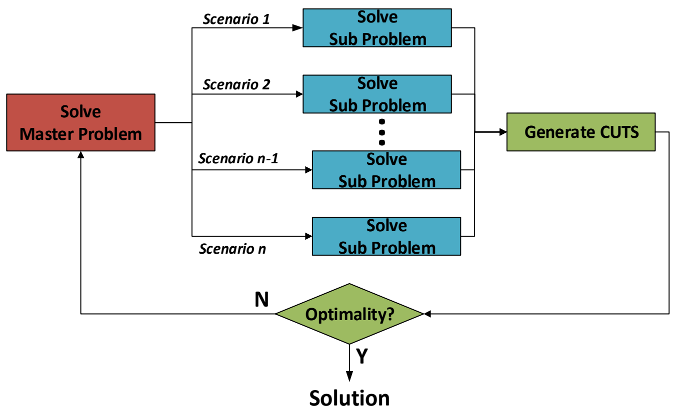


Figure 5-15: Solving Scheme of an L-Shaped Method Application

Next, the basic L-Shaped algorithm is presented following the general guidelines set by Birge and Louveaux (2000):

Standard L-Shaped Method Algorithm:

Step 0: Set $k \leftarrow 0$, s $\leftarrow 0$, $v \leftarrow 0$

Step 1: Set $v \leftarrow v + 1$. Solve the linear program

$$\begin{aligned} Max \quad & z = c^T x + \eta(x) \\ \text{s.t.} \quad & \\ & Ax = b \\ & \eta(x) \geq \alpha_s - \beta_s x \qquad \forall\, l = 1,..,s \qquad \textbf{Eq. 5-25} \\ & x \geq 0 \end{aligned}$$

Let $(x^v, \eta^v(x))$ be an optimal solution. If no constraint (Eq. 5-25) is present, then $\eta_0(x)$ is set to equal $-\infty$ and not considered in the computation of x^v.

Step 2: For $k = 1, \dots, K$ solve the linear program

$$Max \qquad w = q_k^T y$$

$$s.t. \qquad Wy = r_k - T_k x^v$$
$$y \geq 0$$

Let $\boldsymbol{\pi_k}$ be the dual solution associated with the optimality value of each sub-problem k. Define

$$\beta_{s+1} = \sum_{k=1}^{K} p_k * (\boldsymbol{\pi}_{\boldsymbol{k}}^{v})^T T_k \qquad \alpha_{s+1} = \sum_{k=1}^{K} p_k * (\boldsymbol{\pi}_{\boldsymbol{k}}^{v})^T r_k$$

Let $w^v = \alpha_{s+1} - \beta_{s+1} x^v$. If $\eta^v(x) \geq w^v$, stop; x^v is an optimal solution. Otherwise, set $s = s + 1$, add to the constraints set (Eq. 5-25), and return to Step 1.

In this version of the algorithm, Step 1 sets up the solution of the master problem, which is equivalent to the first-stage problem solution in section 0. Note that constraints in Eq. 5-25 are equivalent to those in Eq. 5-9 in the master problem formulation. For the first iteration, however, the value of $\eta_0(x)$ is set to equal $-\infty$ and constraint Eq. 5-9 is an empty set. Step 2 solves each subproblem, which are representative of every scenario that is generated. As it solves each subproblem, it also stores their respective dual solutions. The combination of dual solution vectors is then used to generate a single constraint $\boldsymbol{\eta}(\boldsymbol{x}) \geq \alpha_{s+1} - \beta_{s+1} x^v$. As one can observe, the standard implementation of the L-Shaped method produces a single cut after each iteration, which in some cases may slow its convergence rate. In certain instances one may be able to increase its convergence speed by adding a cut after solving each subproblem instead of waiting until all subproblems have been added (J. Birge and Louveaux 2000). Therefore, each iteration can add K optimality cuts. Next, the multi-cut version of the L-Shaped method algorithm is shown:

Multi-Cut L-Shaped Method Algorithm

Step 0: Set $r = s = v = 0$ and $s_k = 0$ for all $k = 1, \ldots, K$

Step 1: Set v = v+1. Solve the linear program

$$Min \quad z = c^T x + \sum_{k=1}^{K} \eta_k^v(x) \qquad \textbf{Eq. 5-26}$$

$$s.t.$$

$$Ax = b$$

$$\boldsymbol{\eta}(\boldsymbol{x}) \geq \alpha_{l(k)} - \beta_{l(k)} x \quad \forall\, l(k) = 1, .., s_k \qquad \textbf{Eq. 5-27}$$

$$x \geq 0 \quad k = 1, \ldots, K$$

Let $(x^v, \eta_1^v(x), \ldots., \eta_K^v(x))$ be an optimal solution. If no constraint set is present in Eq. 5-27, then $\eta_0(x)$ is set to equal to $-\infty$ and not considered in the computation of x^v.

Step 2: For $k = 1, \ldots, K$ solve the linear program. Let $\boldsymbol{\pi_k}$ be the dual solution associated with problem k. If

$$\eta_k^v(x) < p_k(\boldsymbol{\pi_k^v})(r_k - T_k x^v) \qquad \textbf{Eq. 5-28}$$

Then let

$$\beta_{s+1} = p_k * (\boldsymbol{\pi_k^v})^T T_k \qquad \alpha_{s+1} = p_k * (\boldsymbol{\pi_k^v})^T r_k$$

And set $s_k = s_k + 1$ If Eq. 5-28 does not hold for any $k = 1, \ldots, K$, stop; x^v is an optimal solution. Otherwise, return to Step 1.

In similar fashion as the standard version, the algorithm stores the dual solution vector of each k subproblem. However, in this case, each subproblem generates an individual optimality cut $\eta_k^v \geq \alpha_{s+1} - \beta_{s+1} x^v$. Note that the algorithm checks for optimality at each iteration, as well. Also, since At the end of each iteration, if the optimality conditions are not met, then K optimality cuts are introduced to constraint Eq. 5-27.

5.5 Solving Scheme using Stochastic Cutting Planes

Although the solving scheme created by the standard L-Shaped method (and its accelerated version) are robust to handle the type of problems addressed in this work, it may still be highly inefficient when the number of scenarios is increased. Another reason is that this approach may not work best for purposes of this framework is that its aim is to assess the impact of different scenarios into the optimization framework. In this process, we wish to approximate the number of scenarios needed to capture the set of potential second-stage scenarios by assessing the relationship between the location of the generated scenario in the feasibility space and its impact on the overall optimization solution. Therefore, one of the purposes of this work is the characterization of each entering scenario such that one can assess the degree of similarity an entering scenario and previous problem instances. This will allow the user to predict whether an entering scenario has a higher likelihood of changing the current optimal solution.

A step in that direction is being able to decompose the problem such that scenario samples grow iteratively; for this purpose, a stochastic decomposition approach is sought (Figure 5-16). In this approach, the basic decomposition structure still considers a Bender's approach in which constraints are added iteratively into the first-stage problem formulation. However, instead of solving a pre-defined number of generated sub-problems at once while producing optimality cuts, the algorithm adds a supporting hyperplane defining the master problem's objective function after each iteration. Also, the stochastic decomposition does not solve each iteration to optimality given that each cut is derived using a different number of samples. As the sample size grows, the coefficients of the cutting planes are changed based on a lower bound approximation given by previously generated instances. In this case, since the set of cuts generated from a larger sample size are more stable, the cutting plane coefficients are altered such that cuts generated at later iterations are given more weight in defining the optimal solution (Higle and Sen 1996). In this solving scheme, the algorithm runs until long-term, asymptotical properties show convergence. The only disadvantage of the stochastic decomposition approach is that statistical convergence properties cannot be assured when the vector parameter q^T in the objective function is probabilistic.

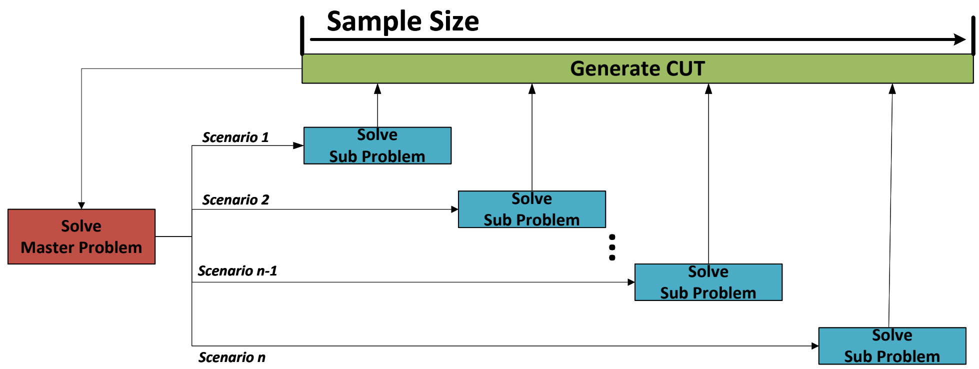


Figure 5-16: Solving Scheme under Stochastic Decomposition

For illustration purposes, a slightly modified version of the two-stage stochastic formulation, shown in Eq. 5-1 and Eq. 5-2, is rewritten next (Eq. 5-29). In this form, parameter vector q^T in the objective function is assumed fix, since convergence cannot be assured otherwise (Higle and Sen 1996). However, it should be noted that for the goal of exploring locally producing systems, being able to incorporate stochasticity in the objective function is not as crucial as the main source of variability in production derives from weather conditions. Furthermore, it is noted that during the beginning exploration stages, prices would be independent from temperature, since it is reasonable to assume that initial production would not affect product offering at the market place.

$$
\begin{aligned}
&\mathit{Max} \quad z = c^T x + E_\omega\left[Q\left(x, \xi(\omega)\right)\right] \\
&\mathit{s.t.} \quad Ax = b \\
&\qquad\quad x \geq 0 \\
&\mathit{where} \\
&Q\left(x, \xi(\omega)\right) = \max_y \{q^T y \mid Wy = h(\omega) - T(\omega)x, y \geq 0\} \\
&\mathit{s.t.} \quad Wy = h(\omega) - T(\omega)x
\end{aligned}
\qquad \textbf{Eq. 5-29}
$$

Next, the general algorithm for a stochastic decomposition approach is described. In this description, the algorithm described by Higle and Sen (1996) is used:

Stochastic Decomposition Algorithm:

Step 0: $k \leftarrow 0$, $V_0 = \emptyset$, $\eta_0 = -\infty$, $x^1 \in X$ is given

Step 1: $k \leftarrow k + 1$. Randomly generate an observation of $\xi(\omega)$, ω^k, independent of any previously generated observations.

Step 2: Determine $\eta_k(x)$, a piecewise linear approximation of $Q_k\left(x, \xi(\omega)\right)$.

a) Solve subproblem (S) and update V_k. Let

$$\pi^k(\omega^k) \in argmax\{\pi(r^k - T^k x^k) \mid \pi W \leq g\},$$

$$V_k \leftarrow V_{k-1} \cup \pi^k(\omega^k).$$

b) Determine the coefficients of kth cutting plane.

$$\alpha_k^k + \beta_k^k x = \tfrac{1}{k}\textstyle\sum_{t=1}^k \pi_t^k (r^t - T^t x)$$

Where$\pi_t^k \in argmax\{\pi(r^t - T^t x^k) \mid \pi \in V_k\}$.

c) Update the coefficients of all previously generated cuts.

$$\alpha_t^k \leftarrow \tfrac{k-1}{k} \alpha_t^{k-1}, \quad \beta_t^k \leftarrow \tfrac{k-1}{k} \beta_t^{k-1}.$$

d) $\eta_k(x) = Max\{\alpha_t^k + \beta_t^k x \mid t = 1, \ldots, k\}$.

Step 3: Solve $Min\ \{cx + \eta_k(x) \mid x \in X\}$ to obtain x^{k+1}. Repeat from Step 1.

The stochastic decomposition algorithm in general can be divided into three main steps after initialization. The first step is generating a random observation ω^k independent from previously generated observations, which assures it maintains its statistical convergence properties (Higle and Sen 1996). The second step is constructing a supporting hyperplane for the linear approximation of $Q_k\big(x, \xi(\omega)\big)$. To do this, we use the set of dual solutions of the second stage problem to construct the kth cutting plane. As additional cuts are added to the first-stage problem, the coefficients previously generated are updated in such a manner to reduce their impact on the objective function value. The combination of cuts generated from the first to the kth cut are then used to delineate $\eta_k(x)$.

5.6 Solving Scheme using Stochastic Cutting Planes while Learning

One of the advantages from taking a stochastic decomposition approach to our problem is that it allows the assessment of the impact of each generated scenario on the convergence of the overall optimization. However, this design still lacks the ability to fully describe the relationship between the generated scenarios and the overall optimization results throughout the solving process. Therefore, as part of this dissertation work, the meta-design for a solving scheme is proposed that seeks to enhance the connection between the characterization of the scenarios inputted and the optimization results. In this case, a combination of dimensionality reduction and machine learning techniques are used to characterize entering scenarios and to predict its impact on future optimization solutions. In this manner, one can gauge the potential impact future generated scenarios will have on the optimization based on previous optimization results. In this section, the general solving scheme is presented followed by a detailed explanation of the feasibility reduction and machine learning components of the proposed approach.

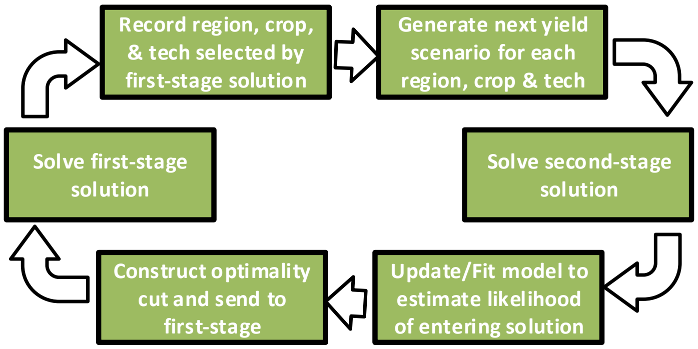


Figure 5-17: Iterative Steps for Stochastic Decomposition with Learning

The solution scheme is comprised of a series of iterative steps, which are illustrated in Figure 5-17. The first step is solving the first-stage optimization problem, which selects the optimal combination of regions, crop, and technologies. This solution is then recorded as inputs to a machine learning component whose purpose is to learn the structure of yield instances of previously selected combinations (i.e. region, crop, and technology combination) from the first-stage solution. As later demonstrated in this section, this allows us to determine the general yield pattern necessary for a crop, region, and technology combination to enter the first-stage solution. The second step in this iterative process is the actual construction of the next yield instance for each crop, region, and technology combination, which are then used as inputs to the second-stage optimization problem in the third step. The fourth step is to update the machine learning model based on previous selections and to predict the likelihood that each of the current yield combinations will change the solution in the next iteration. Finally, in the fifth step, solution of the second-stage problem is used to construct the optimality cut for the next first-stage iteration.

The incorporation of machine learning techniques in characterizing scenarios and their association to first stage solutions is an important contribution of this dissertation work. Although the overall framework is similar to the solving scheme under a regular stochastic decomposition Figure 5-16, in the adjusted framework, a learning level is added to the optimization process that characterizes scenarios that characterizes and predicts upcoming solutions. Specifically, after scenario, a statistical machine learning (SML) model is fitted on previous optimization results and generated instances. The role of the SML model would be to determine the likelihood that a first-stage decision will belong to the optimal first-stage solution. In this case, one can argue that support vector machines do well within this context as they help identify those scenarios that are farther from the boundary between those scenarios that are part of the solutions and those that are not (Platt, 1999). This allows us to gauge the type of previously sampled scenarios that have been included in the first-stage solution. Using the learning capability of machine learning tools, then one can apply sensitivity analysis to estimate the distance in the reduced space of each planting decision from the separating hyperplane that separates the planting decisions included in the first-stage solution. Furthermore, it estimates the probability that an incoming scenario will alter the current first-stage solution based on information derived from previous iterations. The solving scheme under this adjusted framework is presented in Figure 5-18.

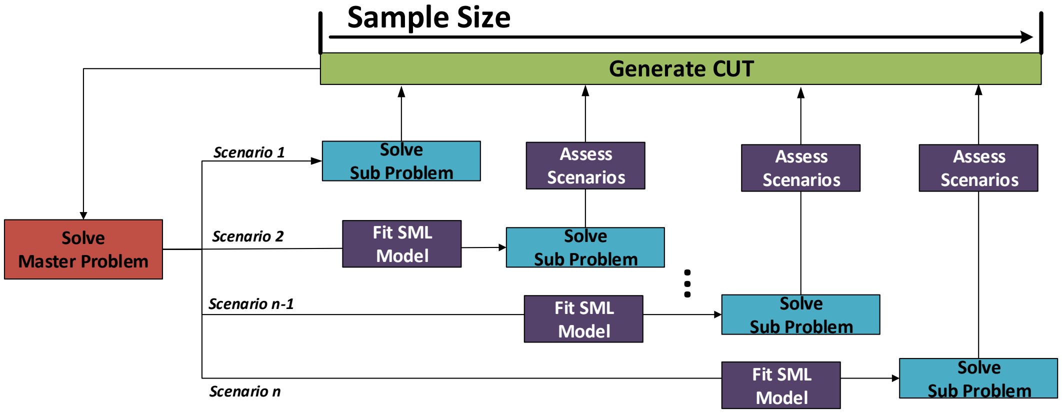


Figure 5-18: Solving under an Adjusted Stochastic Decomposition Algorithm

Next, a modified version of a stochastic decomposition algorithm is presented. In this case, additional intermediate steps are incorporated in the optimization framework.

<u>Modified Stochastic Decomposition Algorithm</u>

Step 0: $k \leftarrow 0$, $V_0 = \emptyset$, $\eta_0 = -\infty$, $x^1 \in X$ is given

Step 1: $k \leftarrow k + 1$. Randomly generate an observation of $\xi(\omega)$, ω^k, independent of any previously generated observations.

Step 2: Determine $\eta_k(x)$, a piecewise linear approximation of $Q_k\big(x, \xi(\omega)\big)$.

a) Solve subproblem (S) and update V_k. Let

$$\pi^k(\omega^k) \in argmax\{\pi(r^k - T^k x^k) \mid \pi W \leq g\},$$

$$V_k \leftarrow V_{k-1} \cup \pi^k(\omega^k).$$

b) Determine the coefficients of kth cutting plane.

$$\alpha_k^k + \beta_k^k x = \tfrac{1}{k}\textstyle\sum_{t=1}^{k} \pi_t^k (r^t - T^t x)$$

Where$\pi_t^k \in argmax\{\pi(r^t - T^t x^k) \mid \pi \in V_k\}$.

c) Update the coefficients of all previously generated cuts.

$$\alpha_t^k \leftarrow \frac{k-1}{k}\alpha_t^{k-1}, \quad \beta_t^k \leftarrow \frac{k-1}{k}\beta_t^{k-1}.$$

d) $\eta_k(x) = Max\{\alpha_t^k + \beta_t^k x \mid t = 1, \dots, k\}$.

Step 3: Transform $T(\omega^k)$ into its principal components $pr_{comp}\left(T(\omega^k)\right)$. Fit support vector model on $d = \{d_t \in pr_{comp}\left(T(\omega^k)\right) | t = 1, \dots, k-1\}$ and dependent variable y^T, the solution for subproblem S_t for $t = 1, \dots, k$-1. Asses the fit of the support vector model on the actual solution for problem S_k

Step 4: Solve $Min\ \{cx + \eta_k(x) \mid x \in X\}$ to obtain x^{k+1}. Repeat from Step 1.

The modified version of the stochastic decomposition algorithm incorporates the machine learning component discussed. Step 0 through Step 2 remain the same as discussed in section 5.5. However, Step 3 transforms ω^k to its principal components with a lower dimension. A support vector regression model is constructed on the transformed set. The data set considered in the construction of this model is based on previous principal components and the solution provided by the optimization model for that instance of the problem. In this manner, the framework seeks to learn the general characteristics of previous solutions and assesses the potential impact that new scenarios will have on the first-stage result.

5.6.1 Characterizing Yield Scenarios through PCA

To facilitate the visualization of the yield instance combinations, as well as to increase the training speed of the learning model component, a series of pre-processing steps are performed in order to reduce its dimensionality (illustrated in Figure 5-19). The first step is to generate temperature scenarios within each of the regions according to the methods outlined in section 5.3.1. From each generated temperature scenario in each region, a single yield instance is outputted for each crop and technology, which can be summarized by a single vector. Each vector is comprised of yield estimates between each planting and harvesting week combinations, which creates a large dimensionality data set. Given the large dimensionality of this yield scenario set, feature reduction techniques can be used, such as principal components, to compress the information in the full data set. A reduced dimensionality set also allows us to better visualize relationships between the different yield scenarios as well as reduce the time to iteratively train the machine learning model. In this section, each of these steps are highlighted and demonstrated in-detail through a practical case study application.

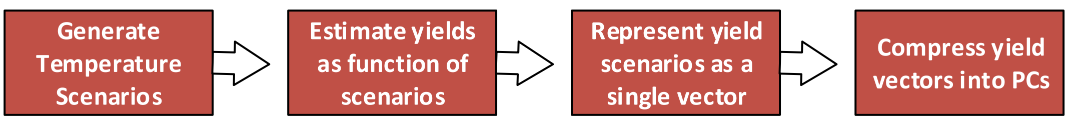


Figure 5-19: Steps to Generating Simple Yield Representations

In a traditional implementation of the stochastic decomposition algorithm, the user can observe changes in the overall optimization solution as new cuts are added to the first-stage problem. These changes in the objective function value are then assessed using statistical-based tests to show convergence (Higle and Sen, 1996). However, no information from the actual location of the scenario in the feasibility space is used to relate it to the first-stage solution. One of the goals from the proposed design is developing techniques through which information gained from large dimensionality scenarios can be used alongside optimization results. In this dissertation, a process is developed that characterizes the relationship between generated scenario instances and the solutions of the first-stage problem. The objective is then to use this information to assess whether generating new scenarios is needed based on previous first-stage solutions. This is done by estimating variations in the first-stage solution vector $\boldsymbol{x}$ as new optimality cuts are iteratively added to the optimization.

Table 5-2: Yearly Yield Vector per Zone, Crop, and Technology

SCEN	WEEKP-WEEKH 1-13	WEEKP-WEEKH 1-13	…	WEEKP-WEEKH 28-52
1	$Yield_{1,iju}^{1,13}$	$Yield_{1,iju}^{1,14}$	…	$Yield_{1,iju}^{28,52}$
2	$Yield_{2,iju}^{1,13}$	$Yield_{2,iju}^{1,14}$	…	$Yield_{1,iju}^{28,52}$
…	…	…	…	,,,
500	$Yield_{500,iju}^{1,13}$	$Yield_{500,iju}^{1,14}$	…	$Yield_{500,iju}^{28,52}$

The first step in this design is facilitating the process through which large-dimensionality scenarios can be represented. To do this, yield scenarios are first flattened into a single vector using the data structure shown in Table 5-2 for 500 scenarios. One should note that for this example, the number of dimensions is fixed to 720, which are the week planted and harvested combinations. Then using dimensionality reduction techniques, the data set of yield scenarios is compressed onto a lower dimension. For this reason, a principal component analysis is implemented, since it is a way to reduce the dimensionality of a dataset by using an orthogonal transformation of the variables in which the resulting vector is an uncorrelated orthogonal basis set (Johnson and Wichern, 1992). This transformation allows us to compress the information contained in the full yield scenario vector into a reduced space accounting for most of the variability in the data set. In this case, since each crop has its own harvest pattern, principal components were constructed per type of crop. From the principal component analysis, it is found that generated harvest yield scenarios can be represented adequately by at most three principal components by type of crop as shown in Figure 5-20. In other words, this means that one can represent each harvest yield scenarios through a three-dimension representation, which greatly facilitates the visualization of crop yield scenarios and their relation to first-stage solutions, as well as speed the training and implementation of an SML model.

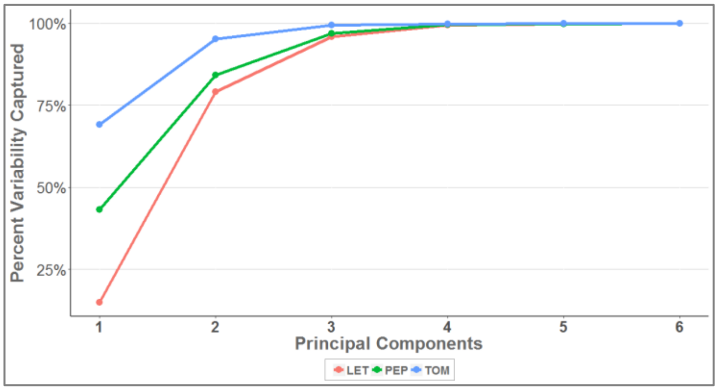


Figure 5-20: Principal Components versus Variance

The principal components can be used to represent the scenario data set by their principal components. For example, Figure 5-21 is a contour plot representation of 500 generated yield scenarios per crop. The figure is decomposed according to crop and the technology, where each zone is represented by a different color. The horizontal and vertical axis are the first and second principal components, respectively, while the contour line is the third principal component. In this figure each point represents a two, principal component representation of yields. In other words, each point captures the estimated yields between any two planting weeks for a single scenario in each region (i.e. a single, full row of Table 5-2). Note that the scenarios under a controlled environment can be represented by single point in each graph given that is assumed that weather patterns do not affect its yields.

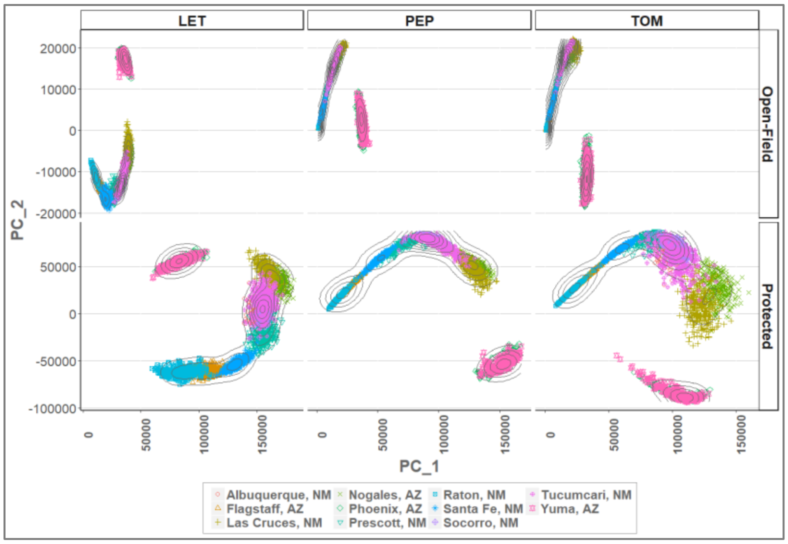


Figure 5-21: Contour Plot of PC3 vs PC1 and PC2

From this figure, one can observe how generated yield scenarios form 5 natural clusters in the reduced space, observable by the 'peaks' of the third principal component contour lines. For example, Phoenix and Yuma share similar yield scenario patterns and form a natural grouping apart from the rest of the regions. Most importantly, from this figure one can also observe how the technologies may alter the yield patterns. Moreover, as it will be discussed in the next section, one can assess the relationship between the location of the yield scenario in the feasible space and how they translate to the first-stage solution vector $\boldsymbol{x}$.

5.6.2 Integrating Scenarios into the Stochastic Decomposition

ONE OF THE GOALS IN THE DEVELOPMENT OF THE STOCHASTIC FRAMEWORK IS CHARACTERIZING SCENARIOS AS THEY ENTER THE OPTIMIZATION. BY REDUCING THE DIMENSIONALITY OF HARVEST YIELD PATTERNS, ONE CAN FACILITATE THE VISUALIZATION OF THE LEARNING PROCEDURE. FOR EXAMPLE, FIGURE 5-22 PRESENTS TWO GENERATED SCENARIOS, IN WHICH THE LINE TYPE (DASHED AND CONTINUOUS) REPRESENT THE DIFFERENT SCENARIOS AND THE COLOR REPRESENTS A DIFFERENT REGION. THE INFORMATION CONTAINED IN THIS GRAPH CAN BE COMPRESSED AND SUMMARIZED BY THREE PRINCIPAL COMPONENTS SHOWN IN FIGURE 5-23. THE LOWER DIMENSIONALITY REPRESENTATION OF ENTERING SCENARIOS AIDS IN VISUALIZING HOW CLOSE EACH YIELD POINT AND SCENARIO IS TO ONE ANOTHER. FURTHERMORE, ONE CAN USE MACHINE LEARNING TECHNIQUES TO LEARN THE RELATIONSHIPS BETWEEN THE LOCATION OF THE YIELD SCENARIOS IN THIS REDUCED FEASIBLE SPACE AND THE FIRST-STAGE PLANTING DECISION ON CROP, ZONE, AND TECHNOLOGY TO USE. BY LEARNING PREVIOUS INPUT-OUTPUT RELATIONSHIPS, ONE CAN USE THIS INFORMATION TO PREDICT EACH PLANTING DECISION'S PROBABILITY OF BELONGING TO THE FIRST-STAGE OPTIMAL SOLUTION DURING THE NEXT ITERATION. FOR DECISION BOUNDARY, ONE CAN USE THIS INFORMATION TO SPACE THAT WOULD IMPROVE ITS PROBABILITY OF ENTERING THE

DETAIL EXPLANATION OF THIS CONCEPT THE READER IS REFERRED TO APPENDIX A

Yield Scenario Characterization. Within this context, one would determine the harvesting pattern needed for a particular planting decision to enter the investment decision.

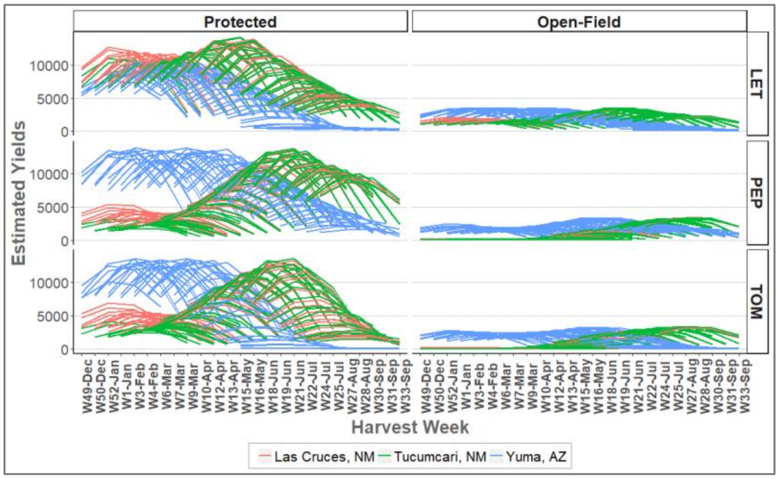


Figure 5-22: Interaction between Different Scenarios and Zones

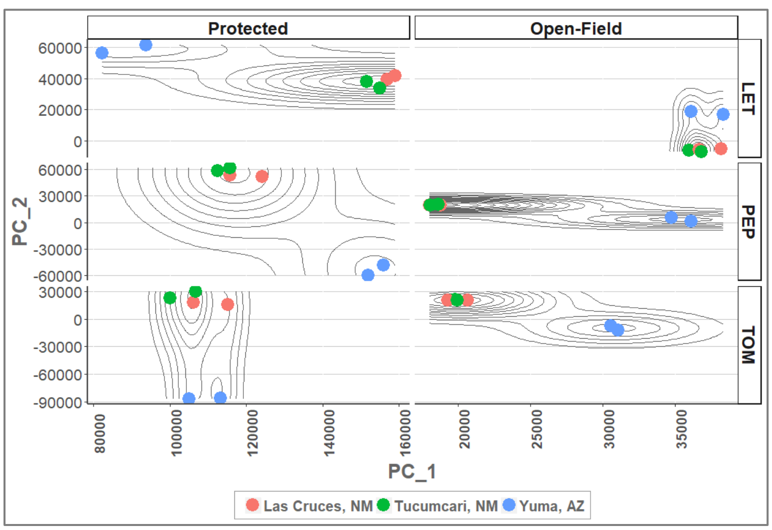


Figure 5-23: Yield Patterns Projected onto PCA Components

Off-the-shelf machine learning techniques can be used to quickly learn the impact previously iterated scenarios have had on optimization results, which can then be used to estimate the probability that the first-stage solution will change. For example, support vector machines can be used to separate the reduced space through the use of hyperplanes to differentiate between those yield scenario points that have entered the first-stage solution and those that have not (Hastie et al., 2009), which is fully explained in APPENDIX A. In this manner, one can characterize the combination of planting decisions most likely to enter the first-stage solution and can provide an additional level of resolution on the relationship between scenarios and the optimization process. Moreover, since support vector machines rely on point classification through the use of these separating hyperplanes, one can transform the distance between each point and the hyperplane to a probability by approximating sigmoid functions to the decision classifiers provided by these models (Bishop, 2006). Points farther away from their closest hyperplane would have a higher probability of being correctly classified than those lying closest to the boundary region. This provides a metric through which the user can decide to terminate the stochastic decomposition algorithm by stopping when the likelihood that a set of planting decisions will belong to the optimal solution. Within the context of our stochastic framework, yield scenarios can be classified based on information derived from previous optimization iterations. After each subproblem is solved, the intermediary statistical learning model is implemented to gauge the probability that each planting decision will be part of the next optimal solution. However, before showing its implementation, we begin with an example of how this process would behave in practice.

Figure 0-24 provides an example of first-stage planting decisions when translated to yield scenarios in their principal component representation. These points have been clustered based on their Ward-distance proximity within the reduced dimensional space and represented by different shapes. In this figure, the size of each point refers to the number of acreage planted for each crop-technology combination, while the color represents the region. The blackened icon refers to those yield scenario points generated during the latest iteration. From this group, those planting decisions with an actual zone name label means they are also part of the first-stage solution during the latest iteration. For example, in this figure, planting lettuce in an open-field environment in Prescott, NM, is part of the first-stage solution during the 20th iteration. Within this graph, additional black points can be observed, which represents the generated yields of planting decisions not selected in the first-stage solution during the 20th iteration. Those points that do not have a label nor have a black marker point are scenarios left-over from previous iterations.

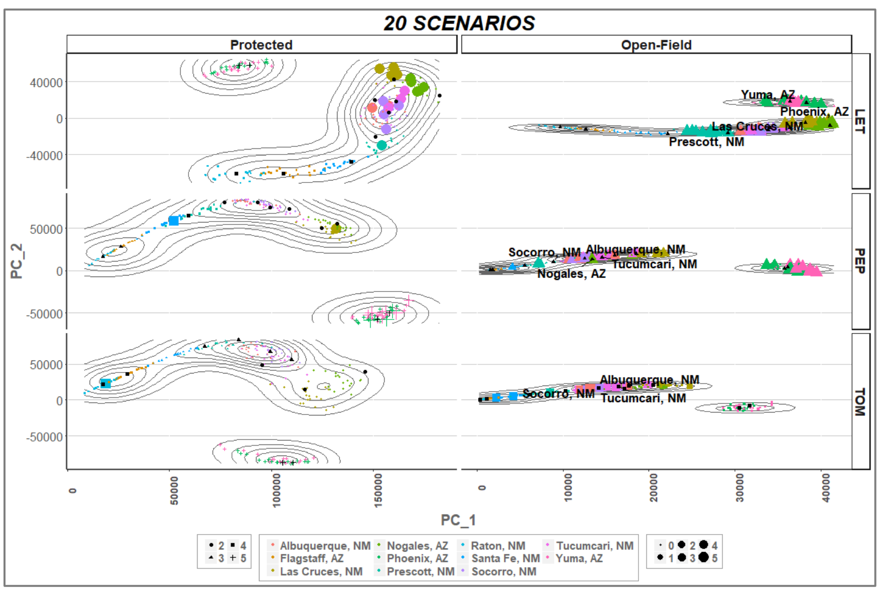


Figure 0-24: First-Stage as Function of Yields after 20 Iterations

As the iterations proceed, one can observe variations in these selections. Figure 0-25 shows the first-stage planting solution after 60 to 100 generated scenarios, from which one can observe a shift in optimal planting selection towards lettuce and peppers under open-field. Again, one can also observe the rest of the generated scenarios that have not been part of the first-stage solution (market by black points). One can also observe how previously generated scenarios span the reduced principal component space. On the right hand-hand side of this figure, one can observe that first-stage solutions have changed little after 100 generated instances. The question is whether one can learn from these changes to identify the relationship between the location of the yield scenarios and optimization solutions and whether one can improve the probability of these planting decisions of entering the first-stage solution. Moreover, these relationships can identify planting decisions that are more stable within the optimal first-stage solutions.

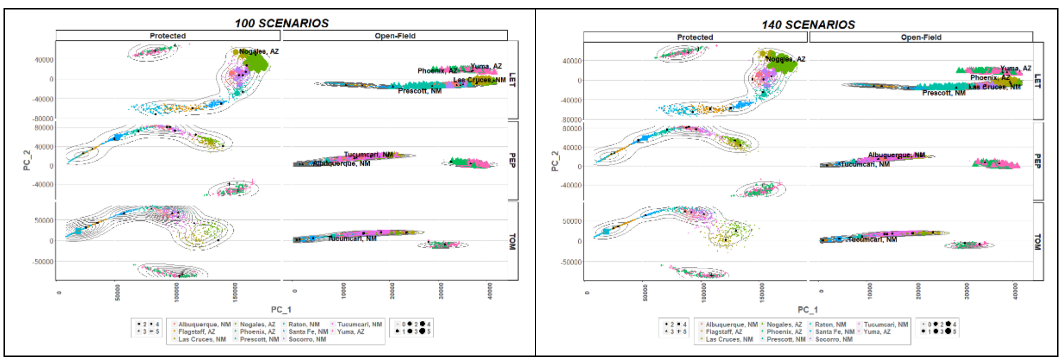


Figure 0-25: First-Stage as Function of Yields after 100 and 140 Scenarios

As mentioned, one of the advantages of reducing the dimensionality of the generated scenario instances is the ability to visualize the relationship to first-stage solutions. Another important reason is that it provides a smaller dimension training set, which can speed up the parameter tuning process of statistical learning models. Furthermore, depending on the type of machine learning model used, one can gain additional insight to the structure of the solution as it relates to the generated scenarios throughout the optimization process. In the case of support vector models, the output is actually the distance between the point and the closest separating classification hyperplane (Hastie et al., 2009). The farther away a point is from the separating hyperplanes, the least likely that point will be misclassified, which in turn can be transformed into probability using a logistic function (Bishop, 2006). Based on this output, one can estimate the probability that a currently outputted optimization solution has been correctly added to the model based on information derived from previous iterations. This provides an additional level of insight to the inner workings of the stochastic decomposition framework, as well as another stopping criterion for the user.

The ability to quickly train and fit a support vector model as the stochastic decomposition scheme progresses allows the user to estimate the probability that a current first-stage solution is correctly classified in first-stage solution based on previous optimization outputs and generated scenarios. For example, using the same graph structure as before, one can assess the probabilities that each first-stage solution has been correctly classified (i.e. previous yield scenario combinations have resulted in the first-stage solution selection). In the left-hand side of Figure 0-26, one can observe the probabilities that each planting decision has been correctly classified to be part of the solution progressively from 20 through 100 scenarios. In this figure, the color-gradient refers to the zone and the probability that each first-stage decision is selected to be part of the first-stage solution in the next optimization iteration. The size of each point refers to the scenario number; the biggest points refers to the latest generated scenario, while the label of the point refers to the zone.

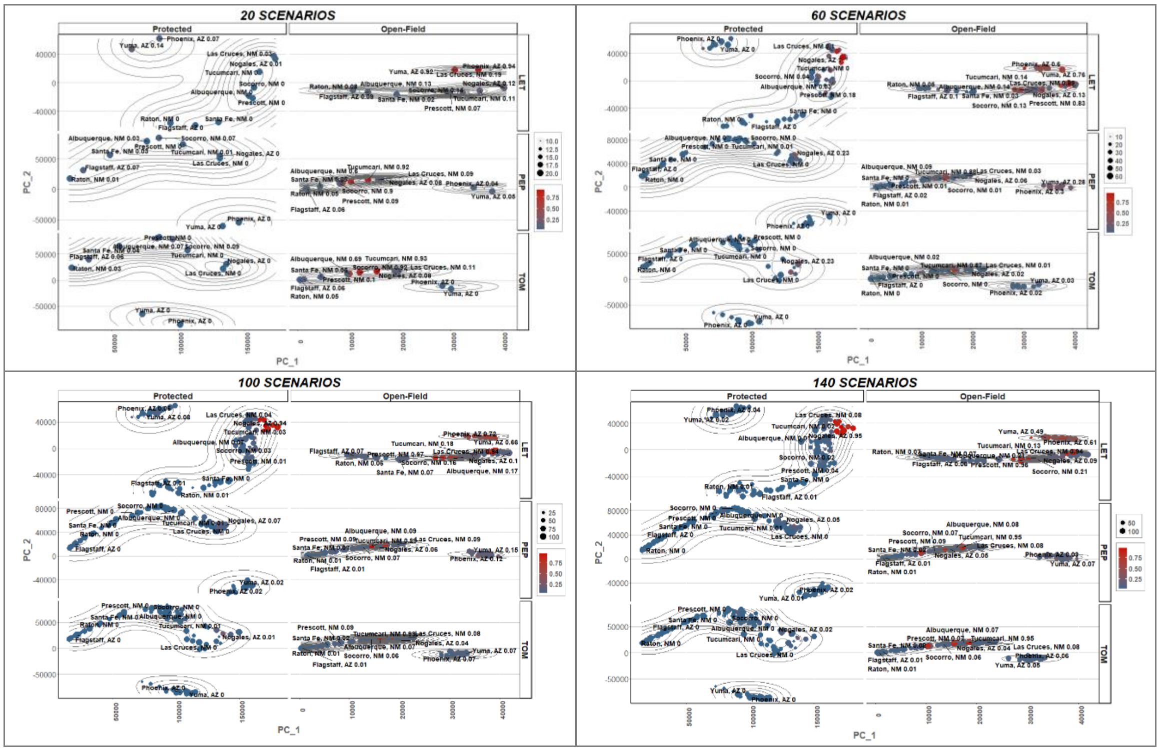


Figure 0-26: Probability of Correctly Classifying First-Stage Solutions

Based on this figure, one can observe that after 20 generated scenarios, the probability for some points is ambiguous. For example, in the case of lettuce production under protected technologies in Tucumcari the probability that this planting decision belongs to the optimal solution is estimated to be 88%, respectively, based on information gained during previous iterations. Similarly, other planting decisions such as in the case of planting pepper under open-field conditions in Phoenix and Yuma have are pushed farther away from the first-stage solution as their probability has dropped as the algorithm progressed. As one continues through the iteration process, one can observe that the planting decisions within the regions selected (stronger red color) slowly increased or decreased to be out of the first-stage solution. This can be observed in Figure 0-27, which shows the average probability for each planting decision of being correctly classified. One should note that this value also includes the probability of being correctly classified when the planting decision is not part of the optimal solution. In this graph, one can observe that the planting decisions for tomato and peppers eventually become more definite, while the indecision with lettuce is visible.

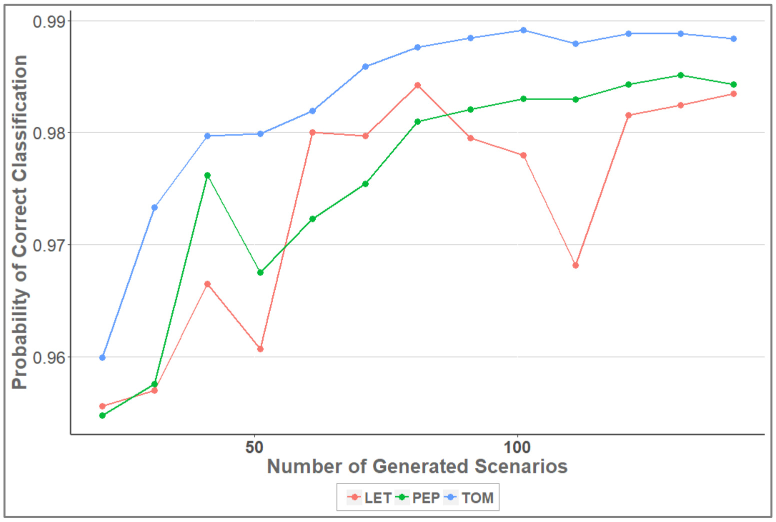


Figure 0-27: SVM Probability of Correct Classification

The question that remains is whether one can leverage from previously generated scenarios and optimization outputs to learn harvesting pattern tomatoes would need to improve its chances of becoming part of the first-stage solution. For example, in the case of pepper production in Las Cruces, one can use the support vector machine developed after 100 iterations to attempt to estimate the probability that a planting decision could enter the first-stage optimization value in the next iteration by varying the location of the yield scenario within the principal component space. This output is shown in the left hand-side of Figure 0-28. In this figure, each black point refers to previous scenario locations within the reduced space. The color gradient refers to the probability that a generated scenario at that location will force the prediction of the planting decision to be part of the optimal first-stage solution. As one can observe, an improved probability is observed when the first principal component is decreased. However, in practical terms, this is not that informative. To gain this perspective, one can reconstruct the harvesting pattern from these principal components and determine the shape of the harvest that would improve the changes of planting tomatoes in Las Cruces from 88%. The reconstructed pattern is shown on the right hand-side of Figure 0-28, in which the last generated scenario is shown in black, while the optimal pattern is shown in red. From a comparison standpoint, one can observe that although the optimal pattern is a bit higher, the yield scenario is relatively similar. This can also be observed by the set of yield scenarios that constantly fall within a high probability area of being included within the first-stage optimization solution.

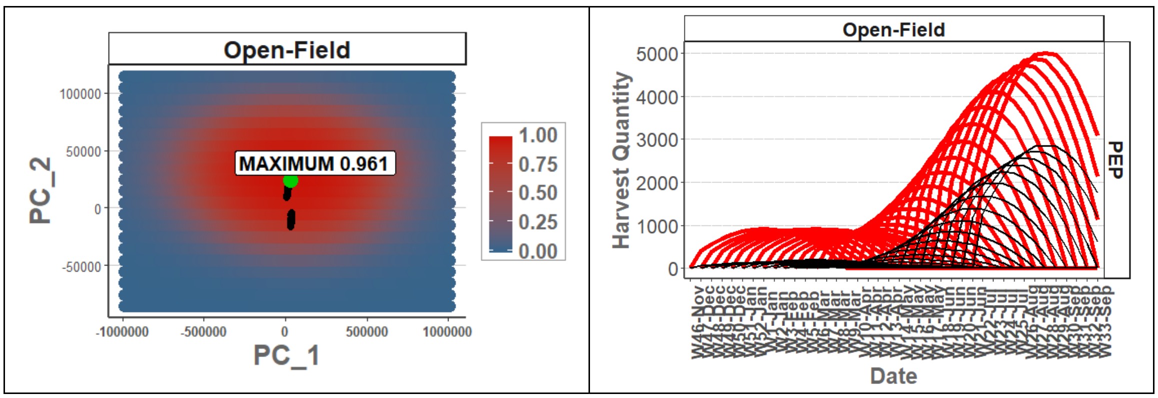


Figure 0-28: Sensitivity of Pepper Harvest PCA in Tucumcari, NM (Protected)

A similar approach is taken to gauge the location of the harvest PCA values for Tucumcari, AZ, which has a very low probability of entering the first-stage solution. From Figure 0-26 one can observe that its probability of entering the first stage solution is consistently below 15%. This is translated in the discrepancy between the level of production required to maximize its probability and the actual expected harvest value. Again, based on a sensitivity analysis on the location of the yield scenario, one could approximate the point at which it would improve the probability that planting pepper in this region will become part of the first-stage solution (left-hand side of Figure 0-29). As one can observe most previous yield scenarios have landed outside the high probability region. However, there is a group of yield scenario points that landed relatively close to this region. From the right-hand side of Figure 0-29, one can observe the higher yield pattern required for lettuce production in Tucumcari that improves its probability of being included in the first-stage solution.

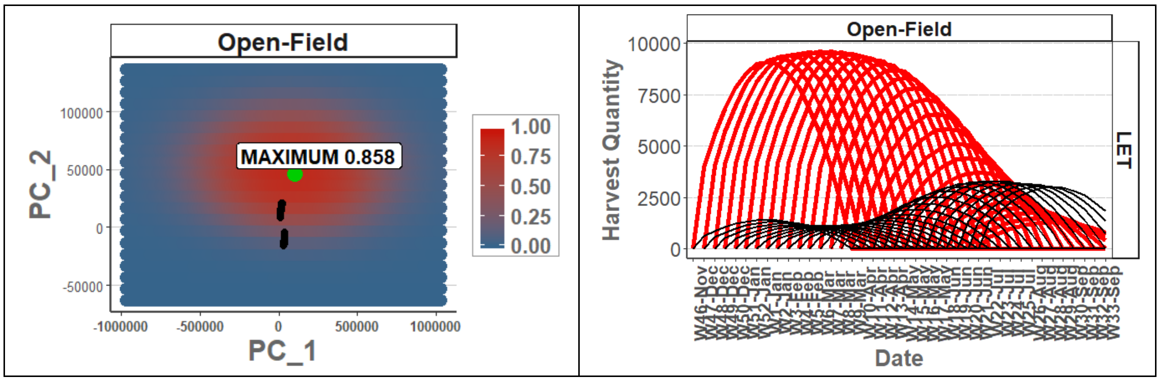


Figure 0-29: Sensitivity of Lettuce Harvest PCA in Tucumcari, AZ (Open-Field)

In this section, one has shown the potential usage that statistical learning tools may have on exploring the effect of the scenarios on an optimization process. Within the context of this problem, the lower dimensionality representation of the harvesting patterns improved the speed of the SML model construction process, which also facilitated the use of a sensitivity analysis on the solution. Nonetheless, it is important to stress that there exist limitations that should be discussed in using theses analysis. The most obvious of these limitations is the fact that the learning procedure is dependent on previously observed inputs and outputs. For example, within the context, the use of greenhouse technologies for tomato production, never entered the first-stage solution, therefore, it is impossible for the modeling procedure to learn variations for this planting. However, it does open the door for future improvements in this process that would seek to learn the influence of low-impact decisions on the optimization procedure by forcing scenarios that would improve the likelihood of being selected in the first-stage optimization solution. A second limitation is the fact that the SML model simply learns the relationships between the scenario inputs and outputs without considering the model formulation or the value of model parameters. This means that exploration of points that are outside of the possibility of actual scenarios may not be reasonable within the context of the actual model itself. However, it does provide some insights to the direction of the scenarios that might result in being part of the first-stage solution.

5.6.3 Advantages of Learning in Stochastic Decomposition

This section is concluded by summarizing additional advantages that could arise from the use of statistical learning techniques within a stochastic decomposition framework. The first advantage of this approach is that it seeks to use available information derived from intermediary steps in the solution process. In a basic implementation of either the L-Shaped method or a stochastic decomposition approach, the information outputted by the solution process is kept within each iteration of the solving scheme and little information is available that would inform the user of how strength of individual components of the solution. This provides the user additional tools to the convergence of the traditional stochastic decomposition scheme. For example, the user can decide to terminate the solving scheme early by assessing changes in the first-stage solution as more scenarios are added to the optimization. Most importantly, the merging between stochastic optimization methods and statistical machine learning tools can have important impact in facilitating the solution process of optimization frameworks. In this implementation, support vector machines were used as they provided an important probability metric important for assessing our results. However, other off-the-shelf predictive tools (both supervised and non-supervised) can be used to characterize the feasible space along the solution process, such as tree-based structures (e.g. random forests and gradient boosting) and neural networks, and facilitate convergence. Other SML models can be used to uncover other type of relationships between master problem solutions and the effect of yield scenarios. Within the overall context of the stochastic framework for the exploration of producing regions, the use of statistical learning tools helps assess the viability of planting decisions that may be close to optimal. For example, based on the analysis of Figure 0-29, one can make different observation afforded by the application of the statistical learning tools, which would otherwise be hidden in the basic implementation. Additionally, one can use these patterns to construct a combination of technology costs and harvest yield necessary within a region to improve its probability of being part of an optimal solution.

5.7 Implementation Results from Solving Schemes

A comparison of the two main solving schemes was performed. For the deterministic, L-shaped method, a clear convergence to an optimal solution was difficult to obtain given the size of the overall problem and the general properties of the solving scheme. Given that the solving scheme utilizes complete information of previous iterations, the solution requires larger memory capabilities when compared to the stochastic version. As shown in this section, the use of a stochastic cutting plan algorithm allows the incorporation of many more scenarios into the overall optimization framework. It also allowed an easier identification of an optimal value. In this section, results from the optimization, as well as their convergence properties, are shown for both types of cutting plane algorithms. One should note that optimization framework was run on an Intel Core i7-6700 3.40GHZ computer with 16.0 GB of memory.

5.7.1 Results from an L-Shaped Method

Using the deterministic cutting plane algorithm, a multi-cut, L-shaped method was first applied to address the stochasticity of this problem. Given the size of the number of variables and constraints for this problem, a balance was attempted between the number of scenarios considered in the optimization and the ability to converge to an optimal solution. For this reason, 20 randomly selected scenarios were used to obtain as many scenarios possible while obtaining a reasonable algorithm convergence. As part of this process, the convergence of aggregated shadow prices is shown to highlight the set of dual variables that had the most problems converging to a stable solution.

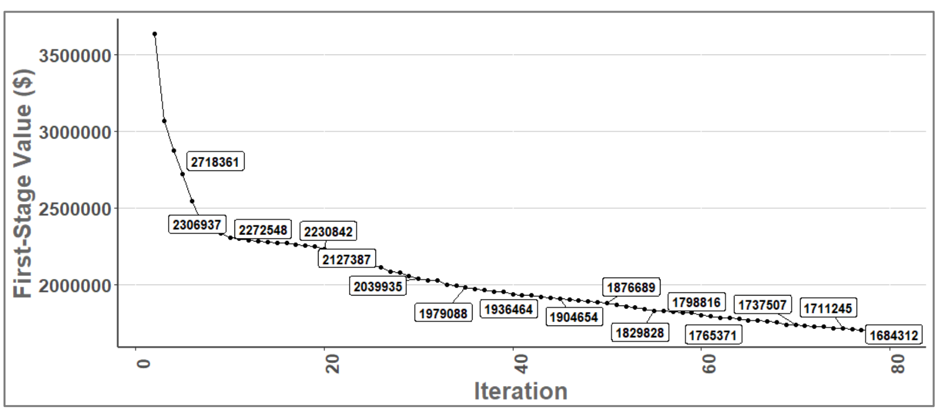


Figure 0-30: Convergence of Multi-Cut L-Shaped Algorithm

Figure 0-32 presents the convergence of the multi-cut, L-shaped algorithm through various iterations of the stochastic optimization. As noted earlier, the convergence of the first-stage solution is not definitely clear. The algorithm was terminated at the point at which the size of the first-stage optimization problem was no longer solvable by the system. In this case, once can observe that the problem seemed to be arriving convergence but had yet to find an equilibrium point. Figure 0-10 presents the aggregated shadow prices for the different dual variables of the second-stage problem's constraints. In this figure, the x-axis corresponds to the cut iteration, while each line represents the different scenario. From this figure, one can observe that the majority of dual variables converged relatively early in the process. However, the dual values for water allocation and warehouse inventory constraints (Eq. 5-11 and Eq. 5-17, respectively, in the stochastic formulation of section 0), have yet to find an equilibrium point for the set of sampled scenarios as the number of iterations progresses. Since these vales directly impact the set of optimality cuts of the first-stage problem (constraint Eq. 5-9 of the master problem formulation), the convergence difficulty is observed in the optimal solution.

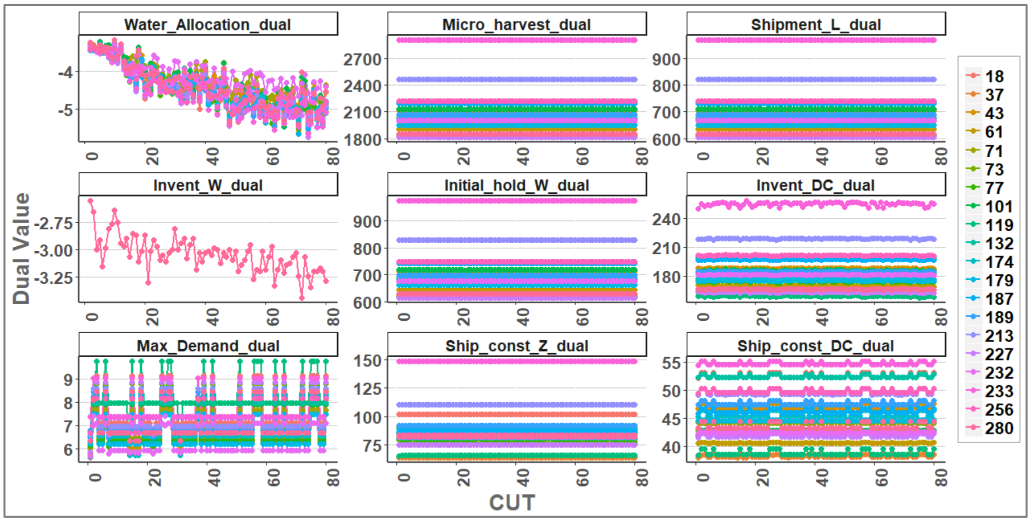


Figure 0-10: Aggregated Shadow Price per Scenario and Cut

Figure 0-32 presents the premature first-stage solution for the stochastic optimization. In this case, once can observe that several first-stage decisions are still transitioning from one solution to another However, one can observe that in general, there is a propensity for the use of protected technologies for the production of lettuce and pepper. On the right-hand side of this figure, one can observe the solution for one of the second-stage scenarios from which one can observe similar complementary production patterns generated under protected technologies.

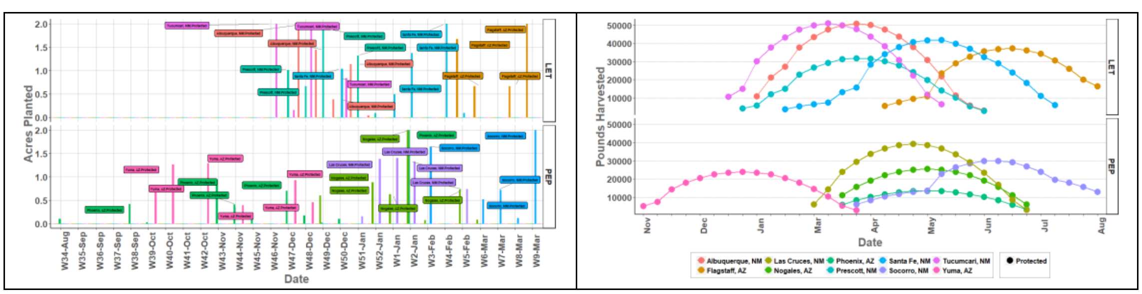


Figure 0-32: Production Strategy using a Multi-Cut, L-Shaped Method Approach

The lack of full convergence of the deterministic cutting plane algorithm exemplifies the need for a more efficient solving scheme. For this reason, the stochastic version of the cutting plane algorithm is needed. A stochastic decomposition framework is also able to consider many more scenarios and find a more stable first-stage solution. Next, the results of the stochastic optimization framework are presented.

5.7.2 Results for Stochastic Decomposition Framework

From the stochastic decomposition framework, one is able to obtain a much more stable production and logistic strategy for the complementary production system. The planting and harvesting patterns under the stochastic decomposition framework is observed in Figure 0-12. As one can observe from this figure, the planting pattern is similar to the results obtained from the deterministic version of the problem. The main product output is pepper in the regions of Las Cruces, Phoenix, and Yuma, while open-field production is expected in Santa Fe. This production is complemented by lettuce production using protected technologies in the rest of the regions, and tomato planting in Santa Fe by the end of the season. Again, one of the important aspects to note in the stochastic version of the optimization framework is the consideration of protected technologies for production. On the other side, one can observe the harvesting patterns that assures year-round product offering within the complementary systems, where pepper and tomato are harvested from protected technologies and tomato under open-field production.

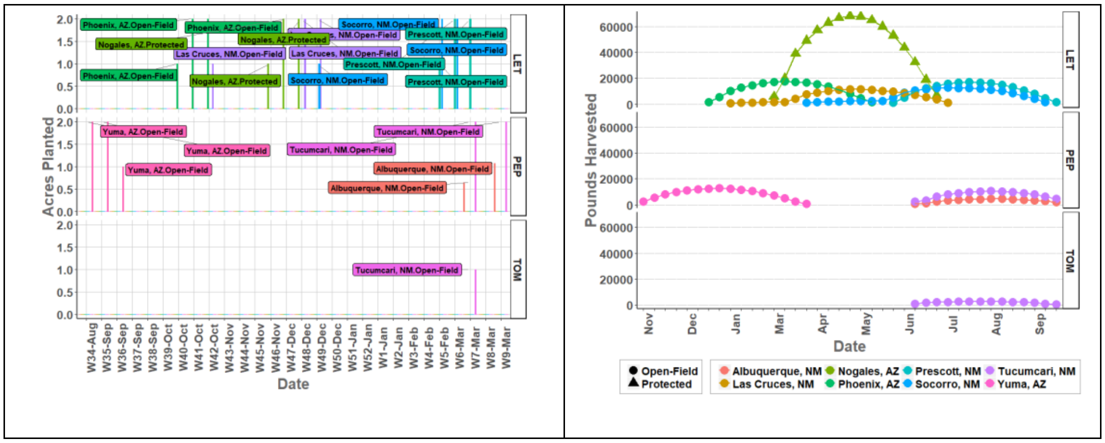


Figure 0-12: Production Strategy using a Stochastic Decomposition Approach

The market shipment strategy is also similar to the one from the deterministic optimization as it can be observed in Figure 0-13. From this figure, one can observe that pepper production is mostly allocated to the Chicago market, while for lettuce production, the main markets are in Pittsburgh, Atlanta, and Philadelphia. In this case, one can also observe that the distribution centers are distributed across Albuquerque, Phoenix, and Tucson. Similarly, for lettuce production, the main market is in Pittsburgh, while other shipments are more evenly distributed across markets in Atlanta, Boston, and Philadelphia. From this figure, one can observe that Albuquerque is the main distribution center from which most is sent to the Pittsburgh market.

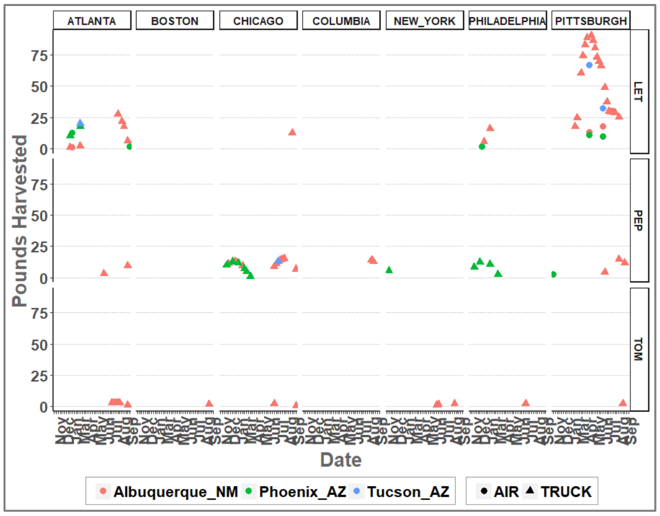


Figure 0-13: Shipment Strategy using a Stochastic Decomposition Approach

The convergence of the stochastic decomposition algorithm seems to converge earlier in the algorithm progress. However, as noted from section 5.6.2, additional information can be obtained from the relationships between yield scenarios and first-stage decisions during the process of convergence. One can also observe that there are subtle changes in the value of the first-stage solution. However, this value oscillates along a fairly stable interval of the optimization value. One should note that at 148 iterations, the size of the problem maxes out the memory of the system. Fortunately, at this point, not only does the value of the first-stage solution is more stable, but the values of the first-stage decision variables are stabilized.

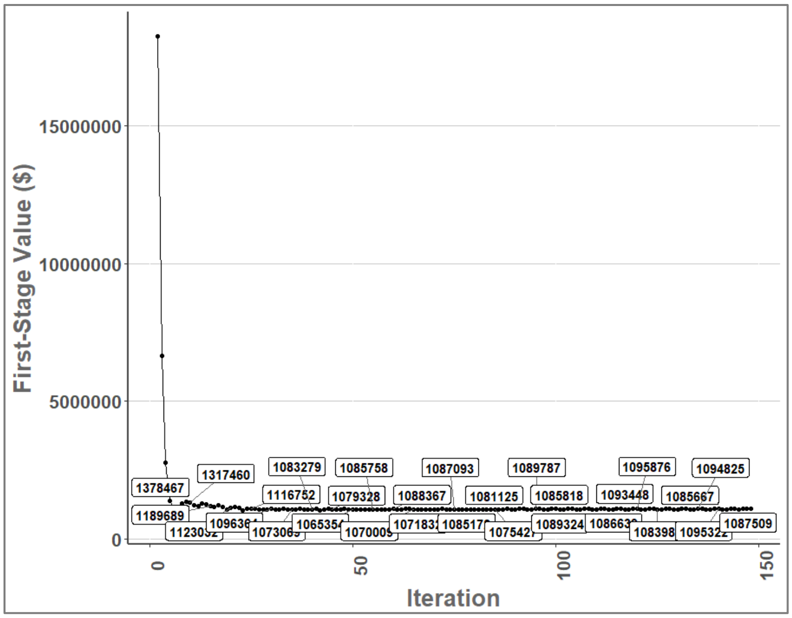


Figure 0-14: Convergence of Stochastic Decomposition Algorithm

Finally, one should note the difference between the first stage value of both stochastic frameworks. In the multi-cut, L-shaped method, the optimization semi-converges to approximately $1.4 million from the whole set of operations using these 20 scenarios. Under the stochastic decomposition framework, the first stage value oscillates roughly around the $1.1 million range. This could be due to several factors including the set of scenarios selected for the multi-cut shaped method, which may be an indication of the information gained from the increased number of scenarios. The other reason might be due to the basic ordering of the scenarios as they enter the optimization. As noted earlier, iterations used at later stages have a greater impact on the solution as the number of samples used to solve the optimization is larger.

5.7.3 Solving Scheme Comparison

As noted earlier, both solution methods are adequate to finding solutions to this problem. In the case of the implementation of the multi-cut L-shaped method provided a solution based on the random selection of 20 samples from the yield spaces. This allows the optimization framework to approximate an optimal solution. However, due to the size of the problem as the number of dual matrices increased, the optimization ultimately was not able to efficiently find and completely convergent solution. This in turn limits the number of scenarios used to approximate a solution. Also since the solutions are continuously added to the first-stage problem in a bundled manner, there is little way of approximating the impact of the individual scenarios on the first-stage solution. A suggested improvement for future expansions of this work would be the improvement of the computational efficiency by developing parallelized framework in which the individual subproblems solutions could be managed more efficiently by separate and individual machines.

For the stochastic decomposition, the algorithm was able to run up to 148 scenarios before the CPU's reaching memory constraint, at which point it had to abandon the optimization process. As one can observe from Figure 0-14, the optimization reached a fairly stable solution early at the beginning of the convergence process. However, there were still some model indecisions regarding individual first stage solutions as shown in Figure 0-27. The important aspect to consider is that as one is solving the stochastic decomposition framework, one is also able to keep track of the stability of the first-stage solution as well. Furthermore, since one is able to use the principal component reduction of the yield patterns, one can also perform a sensitivity analysis on the changes of the first stage solution as a function of the location yields within the sample space. This in turn allows us to reconstruct yield patterns that would be an indication that first-stage decision variable directly related to the yield belonged to the first stage solution.

5.8 Discussion of the Stochastic Framework

This chapter developed a stochastic framework for the exploration of complementary fresh food systems, which addressed some of the shortcomings of the deterministic formulation. As part of its development, contributions were made towards the advancement of agricultural planning research. The first major contribution was in the extension of the deterministic framework through which the variability component of high-value agricultural exploration was implemented. This included an extension to the developed yield functions given sampled yearly temperatures and precipitation values. The second major contribution was the implementation of a stochastic decomposition framework that address some of the limitations observed in traditional multi-cut L-shaped framework in agricultural planning models. As part of this contribution, machine learning techniques are incorporated into the stochastic framework to learn the relationship between first-stage solutions and sampled yield scenarios in the second-stage of the problem.

The expansion of the deterministic framework to consider the stochastic aspects of this problem serves to consider temporal weather variations within different geographical regions. From a practical perspective, it allows investors to determine the impact that yield variations might have on technology selection within a given production zone. From an agricultural planning perspective, it allows the user to construct temperature scenarios to approximate historical weather behavior. The application of feature reduction techniques, such as principal component analysis, serves to not only reduce the dimensionality of the yield set and increase the speed at which the machine learning models are trained, but it also allows the user to visualize the proximity of yield scenarios within the reduced space.

Most importantly, it sets the framework for future refinement of the methodology through which yield scenarios would be constructed.

There are several important advantages to the stochastic decomposition set-up for addressing weather variability as part of an implementation of complementary production system. The first advantage it that allows the user to determine the impact of individual yield scenarios to first-stage solutions as the size of the sample size is increased. In this case, being able to model the relationship between first-stage decisions and yield scenarios allows the user to construct the type of yield scenario that would improves its chances of entering the solution. Moreover, the use of a support vector machine framework allows the user to estimate the probability of entering the first-stage solution as a function of its location within the yield scenario sample space.

Given the scope of the stochastic optimization framework, there are several opportunities for refining and expanding on this work. An area of opportunity for the construction of yield scenarios based on temporal weather conditions is the incorporation of dependency among the temperature vectors for each region. This includes the development of more precise sampling strategies that connects the relationship been climate conditions and physiological requirements of the different crops. Another foreseen expansion to this work is the incorporation of other machine learning techniques that could learn other variable relationships. In this case, machine learning techniques were used to learn the interaction between the first-stage planting decisions and second-stage yield outputs. However, other modeling interactions could potentially be learned. For example, one could gauge whether logistic decisions not directly associated with planting decisions could be determined. In addition, one could develop more exact estimation methods for determining the probability of a generated scenario instance of entering the first-stage solution. Additional considerations that would need to be assessed in future work is the consideration components dictating yield scenarios within each region, such as production and logistic costs.

6. TECHNOLOGY ALLOCATION FOR FARMER INTEGRATION

The deterministic and stochastic framework developed in previous sections take the perspective of a centralized decision-maker in the exploration of potentially producing regions. In this section, an optimization model is developed that considers each farmer as an independent entity. The basic set-up is comprised of an investor seeking a profitable investment in the agricultural development of a region by providing the necessary investments, including protected agriculture technologies, and facilitating labor resource availability to a set of individual farmers. Individual farmers within each region then would decide whether to enter into a production contract with the primary investor based on their own minimum profitability requirements. Thus, the overall framework is constructed as a trade-off between the profitability of the main investment entity and those from individual farmers. One should note that the goal of this problem structure is the construction of the base framework upon which additional coordinating and contract mechanisms and participating supply chain players can be developed and more easily integrated. For purposes of this framework, only the interaction between investor and individual farmers are considered under simple coordinating contracts.

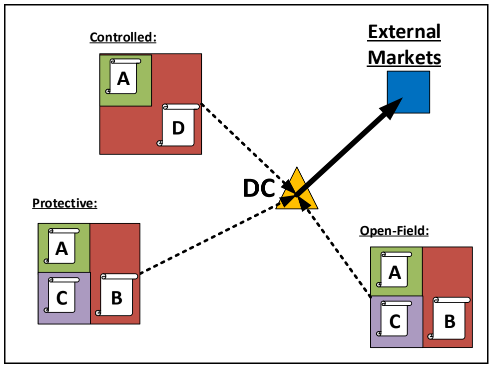


Figure 6-1: Technology Allocation Design

The input to the technology and allocation decision-making model will be the harvesting schedule constructed by the exploration model developed in previous sections. In this case, it is assumed that the investor has contracted the output from the harvesting plan and needs to meet the minimum production requirements. As shown in Figure 6-1, technology space within each of the regions will be available to participating farmers, from which they will select To meet the production requirements, the investor seeks out a group of farmers within an identified region willing to participate given their own profitability expectations. Thus, the overall objective is to optimize the revenues of the investor while also meeting the minimum requirements of the individual farmers (Figure 6-2). Previous works, such as Federgruen et al., 2014) and Huh and Lall (2013) have assessed the problem from a contract allocation problem, in which the coordinating agent provides individual farmers with contract options to satisfy demand requirements. In this contract allocation problems, the investor creates a menu of contracts from which the farmers can select to satisfy their own profit needs and insure that the initial investor will be able to meet his/her own production obligations.

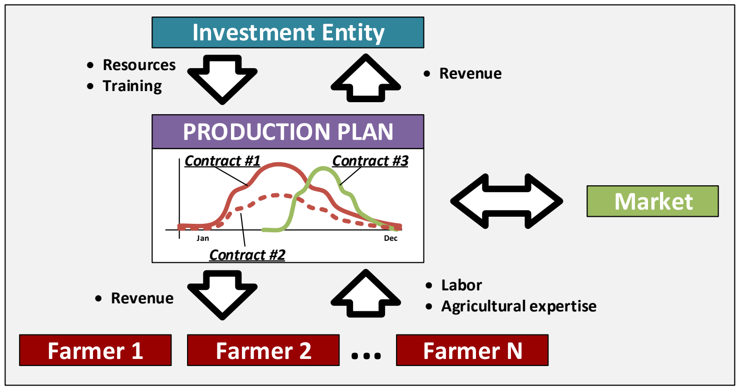


Figure 6-2: Scope of Technology Assignment Framework

One of the main contributions from this component of the dissertation is to develop a framework through which guidelines can be constructed in the allocation of infrastructure and labor resource sharing within a newly producing region. The basic set-up could be in the form of a private investor seeking an attractive return on investment or a government entity seeking to provide the necessary means for the development of locally producing systems. In this case, the external entity would provide the necessary technology and labor resource pools within each selected region in order to meet a previously developed planting and harvesting schedule. This could be in the form of sourcing external workforce through federally mandated programs, such as H-2A geared towards the temporary acquisition of non-immigrant foreign agricultural laborers to supplement local labor pools. A set of simple, production contract alternatives would be made available to farmers within each region to incentivize farming participation. Each contract will be in the form of an assigned acreage space for a particular technology in each region. In return, the individual farmer would receive a certain percentage of the final wholesale market price. One should note that this technology assignment formulation would be the base framework for further expansions into the development of more sophisticated contract mechanisms in future work and provides the basis for exploring different partnership associations between investors and underdeveloped regions.

In the following sections, a deterministic optimization model is constructed that addresses the technology and resource allocation problem. The problem is illustrated through a case study applied to the set of regions illustrated in previous sections. As part of the case study, the different input components to the decision framework are developed. Lastly, the results of this implementation are presented and discussed.

6.1 Technology Allocation and Resource-Sharing Formulation

The objective function of the deterministic optimization model details the profits received by the investing entity in the development of an identified region, which is useful from the perspective of an entity with limited resources seeking to either maximize its profits or the overall benefits of an underdeveloped region. The set of decisions are dictated by the planting and harvesting schedule constructed externally to the formulation. In this set-up, the main investor would be responsible for transporting the harvest from its production source to the external markets. The investor would also be responsible for sharing investment costs of new technology implementations within a new region, as well as assuring that farmers are fully trained in their use. On the other hand, the individual farmer would be responsible for planning labor and production resources in order to meet their contracted obligations. In the case of new technology implementations, individual farmers would be able to eventually own the technologies as long as they participate in the sharing of the investment costs. Additionally, the marginal profits received by individual farmers would need to meet their individual profitability requirements after incurring production and labor costs.

Similar to the optimization model developed in section 4.3, the sets in the technology allocation problem are comprised of the planting and harvesting periods, crops, customers, farmers, and technologies, as well as distribution centers. The sets also include the availability of production contracts. Decision variables in the formulation are comprised of production and logistic decisions that take the production from its source to the consumption market. This includes planting and harvesting, as well as labor resource planning, decisions. Each individual farmer would also be tasked with choosing their own level of participation. Finally, the set of logistic decisions such as shipment and inventory movements are also considered in similar fashion to previous formulations in section 4.3.

Sets:

$t_p \in T_p \subset T$:	Set of planting periods in T
$t_h \in T_h \subset T$:	Set of harvesting periods in T
$j \in J$:	Set of crops
$c \in C$:	Set of customers
$u \in U$:	Set of technologies
$z \in Z$:	Set of zones
$d \in D$:	Set of distribution centers
$f \in F(z)$:	Set of farmers per zone z
$r \in R$:	Set of contracts

Decision Variables:

$DConSel_{rfju}$:	1 if contract r is selected by farmer f for crop j and tech u; 0 o/w
$DPlant_{jfruz}^{t_p}$:	Acres planted of crop j under tech u by farmer f at time t_p under contract r
$DMicroH_{jfruz}^{t_h}$:	Harvest of crop j by farmer f under tech u at time t_h under contract r
$DOPL_f^t$:	Number of seasonal laborers hired by farmer f at time t

$DOPT_f^t$: Number of temporary laborers hired by farmer f at time t

$DHire_f^t$: Number of laborers hired by farmer f at time t

$DFire_f^t$: Number of laborers fired by farmer f at time t

$SLZ_{jfz}^{t_h,t}$: Qty. shipped of crop j from farmer m *to* region z at time t harvested at t_h

$SZD_{jzdm}^{t_h,t}$: Qty. shipped of crop j from region z *to* DC d at time t harvested at t_h on mode m

$SDC_{jdcm}^{t_h,t}$: Qty. shipped of crop j from DC d *to* market c at time t harvested at t_h on mode m

$\mathrm{Invw}_{\mathrm{jz}}^{\mathrm{t_h}}$: Inventory of crop j at zone z at time $\mathrm{t_h}$

$\mathrm{Invw}_{\mathrm{jz}}^{\mathrm{t_h}}$: Inventory of crop j at zone z at time $\mathrm{t_h}$

The parameter set is comprised first of the number acres available to all farmers within each region, which is set by the original planting and harvesting schedule. As part of this set are the different fixed contract configurations offered to the farmers at the beginning of the planting season. This includes the number of acres to be planted by farmers under different contracts and the percentage of the final selling price received. For example, each contract is assigned a particular land portion of the technology to plant an assigned crop. In this case, if the farmer is willing to plant more acreage, he/she can decide to contract a larger portion of the technology and have higher harvest requirements during the harvesting period for which they will be rewarded by a higher portion of the market sales received. However, farmers would also have the option of planting more acreage at the lower contracted price, which would allow him/her to increase revenues while limiting his/her contracted obligation. In this case, their willingness to participate in a given contract is based on their own minimum profit value. In future work, this willingness would also be dictated by the level of risk in production. It should also be noted that the design of individual contracts is constructed outside of the optimization framework. Incorporated to the set of parameters are the estimated number of labor resources needed for each crop for planting and harvesting. Finally, the set of production and logistic costs are also considered, which dictate logistic decisions.

Parameters:

$LAvail_{uz}$:	Land available for technology u in zone z
$Yield_{jzu}^{t_p}$:	Yld of crop j planted in t_p using tech u
$CPrice_{jru}$:	Price percentage received for crop j under contract r using tech u
$MinProd_{jz}^{t_h}$:	Minimum production required for crop j at time t_h in zone z per acre
$PMinUtil_f$:	Minimum profitability required by farmer f
$CONPlant_{rju}$:	Contracted acres to plant of crop j using tech u under contract r
$PLaborH_j^{t_pt}$	Number of laborers required for harvesting crop j between t_p and t
$PLaborP_j^{t_pt}$	Number of laborers required for harvesting crop j between t_p and t
$MFix_z$	Number of laborers available in zone z
$Ctech_u$:	Amortized investment cost of technology u
$Cplant_j$:	Cost of planting a full acre of crop j
$Coper_u$:	Cost of operating technology u for one year
$CTLZ_z$:	Transportation cost from region z warehouse
$CTZD_{zd}$:	Transportation cost from region z to DC d
$CTDC_{dcm}$:	Transportation cost from DC d to customer c using transportation mode m
$Ccase_j$:	Packaging cost for crop j
$Cwater_z$:	Cost of additional water allocated to region z
$Cavail_f$:	Capital available to farmer f
LT_{dc}:	Transportation time between DC d and customer c

As mentioned, the objective function of the optimization is the profits received by the main investor, as he/she sells the product at the market place after incurring supply chain costs from taking the product from its production source to the market place, as well as payment to farmers within each region based on their selected contract. In this case, it is assumed that the same amortized investment costs as those in section 4.4.2 would be incurred. The investor would also be responsible in the cost-sharing of new technology implementations with individual farmers.

Maximize:

$$\begin{aligned}&\sum_{t_h jcmfrz} Price_{jc}^{t_h} * SDC_{jdcm}^{t_h,t} \\ &- \sum_{t_h jcfrz} Price_{jc}^{t_h} * CPrice_{jr} * DMicroH_{jfruz}^{t_h} \\ &- \sum_{jqzt_h t} Invw_{jz}^{t_h t} * Cw_z - \sum_{jqdt_h t} Invd_{jd}^{t_h t} * Cd_d \\ &- \sum_{z,t_h,j} Pack_{z,j}^{t_h} * Ccase_j \\ &- \sum_{jdmt_h tc} SDC_{jdcm}^{t_h,t} * CTDC_{dcm} - \sum_{jzt_h td} SZD_{jzdm}^{t_h,t} * CTZD_{zd} \\ &- \sum_{t_p jfuz:f\in F(z)} DPlant_{jfruz}^{t_p} * Ctech_{uz}\end{aligned} \qquad \textbf{Eq. 6-1}$$

Subject to:

$$\sum_{t_p} DPlant_{jfruz}^{t_p} \le LAvail_{uz} \qquad \forall u \in U, z \in Z \qquad \textbf{Eq. 6-2}$$

$$\sum_{rju} DConSel_{rfju} \le 1 \qquad \forall f \in F(z) \qquad \textbf{Eq. 6-3}$$

$$\sum_{t_p} DPlant_{jfruz}^{t_p} \ge CONPlant_{rju} * DConSel_{rfju} \qquad \forall f \in F(z), \forall j \in J \qquad \textbf{Eq. 6-4}$$

$$DMicroH_{jfruz}^{t_h} \le \sum_{t_p} DPlant_{jfruz}^{t_p} * YDist_j^{t_p t_h} * Yield_{j_p}^{t} \qquad \forall t_h \in T_h, \forall r \in R, f, z \qquad \textbf{Eq. 6-5}$$

$$\sum_{fr} DMicroH_{jfr}^{t_h} \ge MinProd_{jz}^{t_h} \qquad \forall t_h, j, z\ u \qquad \textbf{Eq. 6-6}$$

$$DOPL_f^t + DOPT_f^t \ge \sum_{t_p t_h juz} DPlant_{jfruz}^{t_p} * \left[PLaborP_j^{t_p t} + PLaborH_j^{t_p t}\right] \qquad \forall t, f \qquad \textbf{Eq. 6-7}$$

$$DHire_f^t - DFire_f^t = DOPL_f^t - DOPL_f^{t-1} \qquad \forall t, f \qquad \textbf{Eq. 6-8}$$

$$\sum_{f} DOPT_f^t \leq PMTemp_z \quad \forall t,z \qquad \textbf{Eq. 6-9}$$

$$\sum_{t,f} DHire_f^t \leq MFix_z \quad \forall z \qquad \textbf{Eq. 6-10}$$

$$\sum_{t} \left(OPL_f^t + OPT_f^t + Fire_f^t + Hire_f^t\right) \leq MFix * \sum_{rju} DConSel_{rfju} \quad \forall f \qquad \textbf{Eq. 6-11}$$

$$\sum_{t_h,j,r,c} PCPrice_{jr} * DMicroH_{jfr}^{t_p} - \sum_{t_p rjfu} DPlant_{jfruz}^{t_p} * CPlant_j - \sum_{t} \left(OPL_f^t * Clabor + OPT_f^t * CTemp + Hire_f^t * Chire\right) - \sum_{t_p t_h z} WA_z^{t_p t_h} * Cwater_z - \sum_{jft_h tz:f\in F(z)} SLZ_{jfz}^{t_h,t} * CTLZ_z \quad \forall f \qquad \textbf{Eq. 6-12}$$

$$\sum_{t_p juz} DPlant_{jfruz}^{t_p} * Ctech_{uz} \geq \sum_{rju} PMinUtil_f * DConSel_{rfju}$$

$$SLZ_{jfz}^{t_h,t} = \sum_{ur:q=2,f\in F(z)} DMicroH_{jfruz}^{t_h} \quad \forall t_h, j, z, f \in F(z) \qquad \textbf{Eq. 6-13}$$

$$Pack_{jz}^{t_h} = \sum_{fu:q=2,f\in F(z)} MicroHarv_{jfu}^{t_h} / ContCap_j \quad \forall t_h, j, z \qquad \textbf{Eq. 6-14}$$

$$\sum_{jq:q=2} Pack_{jz}^{t_h} \leq WZCap_z + AddWCap_z \quad \forall z \qquad \textbf{Eq. 6-15}$$

$$Invw_{jz}^{t_h t_h} = \sum_{f\in F(z)} SLZ_{jfz}^{t_h t_h} \quad \forall t_h, j, z \qquad \textbf{Eq. 6-16}$$

$$Invw_{jz}^{t_h t} = Invw_{jz}^{t_h t-1} + \sum_{f\in F(z)} SLZ_{jfz}^{t_h t} - \sum_{d} SZD_{jzdm}^{t_h,t} \quad \forall t_h, t, j, t > t_h \qquad \textbf{Eq. 6-17}$$

$$Invd_{jd}^{t_h t_h} = \sum_{z} SZD_{jzdm}^{t_h,t} - \sum_{cm} SDC_{jdcm}^{t_h,t} \quad \forall t_h, j, d \qquad \textbf{Eq. 6-18}$$

$$Invd_{jd}^{t_h t} = Invd_{jd}^{t_h t-1} + \sum_{z} SZD_{jzdm}^{t_h,t} - \sum_{cm} SDC_{jdcm}^{t_h,t_h+LT_{dc}} \quad \forall t_h, t, j, d, \; t > t_h \quad \textbf{Eq. 6-19}$$

The optimization is first constrained by the number of acres assigned to each location and technology based on the output from the optimal planting and harvesting schedule (Eq. 6-2). Eq. 6-3 limits the farmer to enter into one seasonal contract. Eq. 6-5 constraints the number of acres planted by the amount selected in the contract. Eq. 6-5 and Eq. 6-6 keep track of the harvested amount, as well as ensure that the minimum production level estimated by the harvesting schedule. Eq. 6-7 assures that the number of laborers at hand satisfy the planting and harvesting requirements across different periods. Eq. 6-8 assures the balance of laborers hired and fired with respect to the actual number of people working in the fields across time. Eq. 6-9 and Eq. 6-10 constraints the number of available laborers to the temporary and fixed labor pools within each of the zones, respectively, while Eq. 6-11 forces the workers available to a farmer if he/she has selected the contract. Eq. 6-12 assures that the minimum utility of each farmer is satisfied after considering for the production and logistic costs incurred from production. In this case, it is assumed that the individual farmers would also be involved in the cost-sharing of more sophisticated technology implementations. Finally, Eq. 6-13 to Eq. 6-19 detail the movement from production through the created logistic network to the market place similar to Eq. 4-9 to Eq. 4-15 from section 4.3.

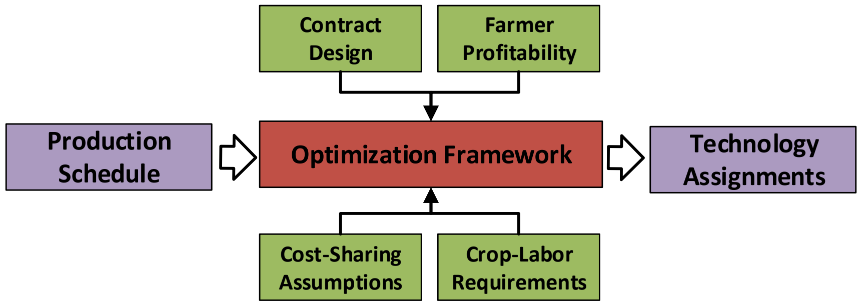


Figure 6-3: Overall Optimization Design

The overall optimization design is presented in Figure 6-3. The main input to the optimization framework is the production schedule derived from the exploration and identification components in chapter 4 and 5. There are also four main external components to this framework. The first component is the design of the individual contracts available to the individual farmers. For purposes of an initial implementation, simple contract mechanisms are constructed. The second component to this framework is an approximation of minimum profitability requirements available within the different regions. In this case, these minimum profitability requirements will be constructed according to randomly generated values that seeks to capture different types of potential farmers. The third component are the type of cost-sharing agreements that could be constructed between the overall investor and the individual farmers. This would include, for example, the percentage of the investment costs that farmers would be willing to provide in return for being able to keep the technology. The final component in this framework is the estimation of labor requirements needed for the different types of crops. For this component, historical production information is used to generalize the number of laborers needed throughout the production cycle of the different crops. Finally, one should stress that each of these individual components can easily be refined outside of the optimization formulation.

6.2 Case Study of Technology/Resource Sharing Framework

To demonstrate the application of this framework, previously developed case studies in chapters 4 and 5 are expanded to consider the technology assignment and labor planning components. As part of the development of the case study, labor planning and harvesting requirements per crop are estimated based on available historical information generalized from producing regions in Yuma, AZ, and Sinaloa Mexico. This labor requirements are then inputted to the technology and resource planning formulation. Also, minimum profitability requirements are synthetically constructed for farmers within each region, as well as a set of potential contract alternatives provided to farmers. As noted earlier, the set of contracts are dependent on the amount of acreage assigned to each individual farmer and the percentage of the final sale price they would receive. Finally, planting and harvesting schedules outputted in section 4.6 are used as inputs to the current optimization case study for a small problem instance. In this case-study, it is assumed that each region will have 20 potential farmers available for participation.

6.2.1 Estimation of Labor Resource Needs

To estimate the number of laborers required per crop within each region, historical production schedules from Yuma, AZ, and Sinaloa Mexico were used for romaine lettuce, bell peppers, and tomatoes. Figure 6-4 presents the number of laborers required per acre based on historical production plans for a single production season for each individual crop. For example, in the case bell pepper, the number of laborers at the beginning of the production season would require approximately 2-to-3 workers when compared to those towards the end at between 1 and 2. In this case, the goal is to generalize labor requirements for each crop to the different producing regions, so that they can more easily be incorporated into the optimization formulation through parameter $PLaborP_j^{t_p t_h}$, the number of laborers required for planting activities of crop j between planting week t_p and harvesting t_h. To achieve this goal, support vector machine models were used to learn the interaction between planting and harvesting weeks and the number of laborers required per crop. This also allow the smoothing and extrapolation of labor needs per acre within other regions for the set of crops considered.

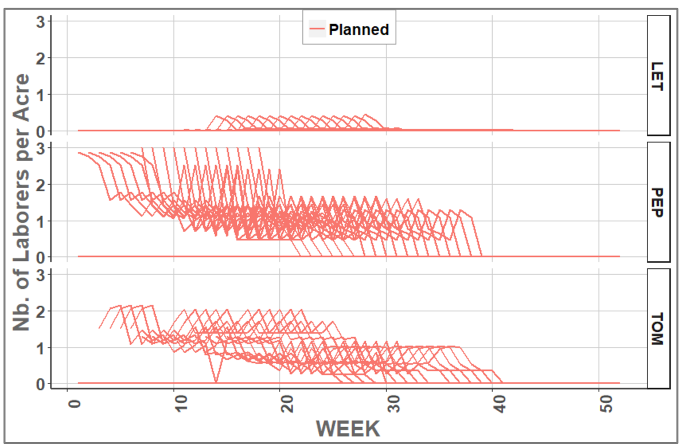


Figure 6-4: Est. Number of Laborers per Week using Planned Schedule

Using historical production plans for each crop within each region, support vector machines were used to train a prediction model to learn the relationship between the number of laborers needed in production and the planting/harvesting week, type of labor, and crop. In this case, the number of laborers is used as a dependent variable, while the planting and harvesting weeks, as well as the type of crop and labor are used as independent variables (Eq. 6-20).

$$Nb\ Laborers \sim t_p + t_h + Crop + Labor_{Type} \qquad \textbf{Eq. 6-20}$$

The ability to learn this relationship also helps generalize and smooth labor requirement estimates for each crop within each region. One should note that more precise labor requirement estimations could be applied. However, for purposes of this simple case study, the smoothing and extrapolation capabilities provided by this generic approach meets our implementation needs. Furthermore, this refinement could be performed outside of the optimization formulation as more data is collected and easily be tuned and incorporated.

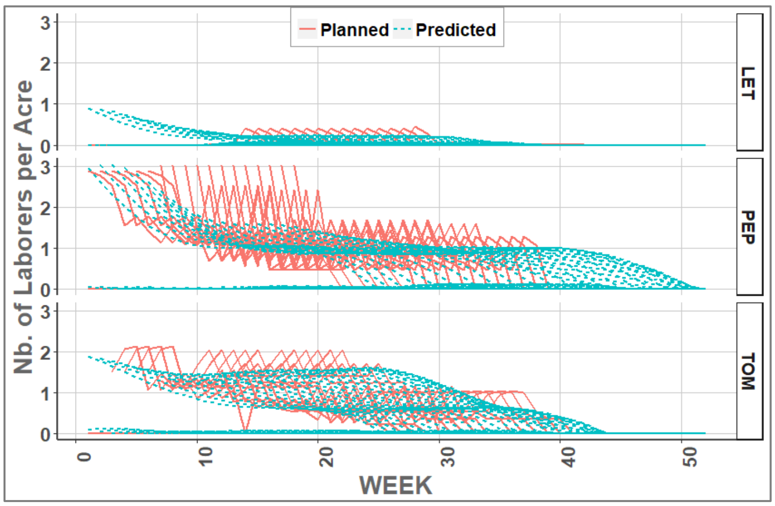


Figure 6-5: Est. Number of Laborers per Week (Smoothed using Prediction)

Figure 6-5 presents the output from the learning model across different production weeks. As one can observe, the predicted output overlays the actual planning numbers, while also serves to smooth the original values. Using this prediction model, one can generalize labor requirements within other regions that have decided to produce each of these crops. The outputs from the prediction model can then be used as inputs to Eq. 6-7 of the modeling formulation, which also serves to plan labor resource use by each farmer.

Table 6-1: Additional Labor Assumptions within all Regions

Labor Type	Availability	Labor Cost	Hire Cost
Full Time	2	$1600	$2000
Temporary	25	$800	$2000

The number of laborers used within each region also depends on the availability of full-time and/or temporary labor pools. Table 6-1 presents the availability of labor pools assumed within the different regions. In this case, it is assumed that there are 2 full-time seasonal laborers per region and approximately 25 temporary laborers given that that we are considering small acreage farming systems. The cost per week for a temporary laborer is assumed to be set at $10 an hour, which amounts to $800 per week. The cost for a full-time laborer is assumed to be set $20 an hour, given that additional benefits would be provided. Finally, it assumed that the cost to hire a new worker onto the operation will be approximately $2000, which would include the process of recruiting new laborers and providing necessary training and equipment.

6.2.2 Contract Construction/Farmer Profitability Requirements

The next component in this framework is the construction of simple contracts made available to farmers within each region with the purpose of incentivizing participation in agricultural production. In this case, the set of contracts were based on a combination of the number of acres planted and the percentage of the final sales price received by the farmer. The design of the contract was such that as the farmer decided to increase the number of acres to plant, the farmer would be compensated by a higher percentage of the final price. Conversely, if the farmer decides to plant less acres of land, the percentage received by the farmer would be reduced. Therefore, each farmer selects the contract that best meets his/her needs based on his/her minimum profit requirements. One should note that the literature related to contract design and different manners of constructions is well studied and it is highly likely that more efficient mechanisms exists. However, as mentioned earlier, the focus is on the construction of the framework as it relates to the development of local agricultural production.

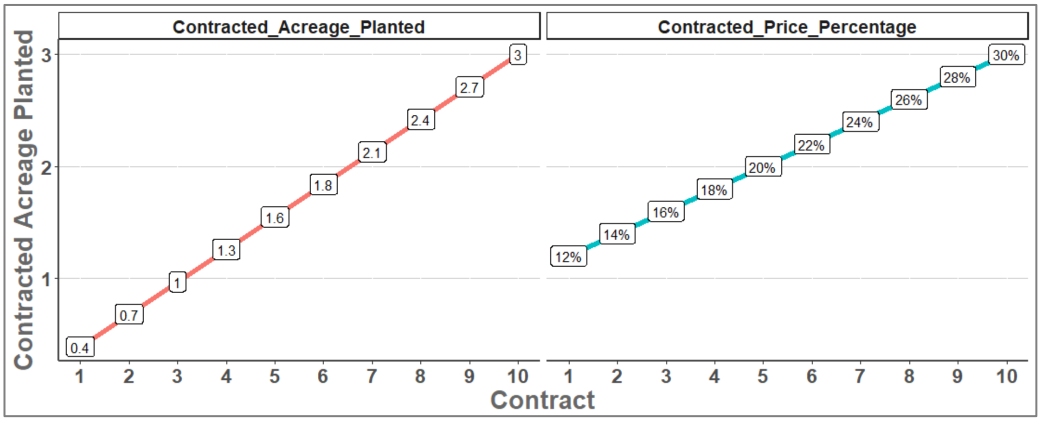


Figure 6-6: Set of Contracts Available to Farmers

Figure 6-6 presents the set of contracts available to farmers within the different regions. In this case, the number of acres is limited to 3 acres at which the farmer is able to keep 30% of the total sale price of the wholesale market value, since he/she is willing to take on higher production values. Conversely, the lowest number of acreage that a participating farmer can take is 0.4 acreage at which he/she would receive only 12% of the total wholesale market price. Thus, each farmer would select the contract that meets his/her own profit requirements. If for any of the farmers, the set of available contracts does not meet his/her minimum profit requirements, he/she simply does not participate in production. One should note that the individual farmer would also be able to commit to more than the production assigned as long as the overall acreage availability within each region is not surpassed. At this point, it is important to note that the investor will be responsible for investing and training farmers in the use of new production technologies, as well as transporting the product from its production source to the end consumer market. The individual farmer would be responsible for production and labor management, as well as cost-sharing for investments in protected and controlled technologies.

The next component in this framework is constructing the set of profitability requirements by farmers within each region. In this case, we assume that the profits required by each farmer per acre are distributed lognormally ($\sim lognormal(0,1)$) and then multiplied by a factor of $1,000. This allows multiple samples of minimum farmer's profitability requirements. One should note that other methods of eliciting farmers' minimum profitability exists that adhere more strictly to more realistic values. However, as before, these refinements could be performed outside of the formulation and be continuously updated through different solution iterations.

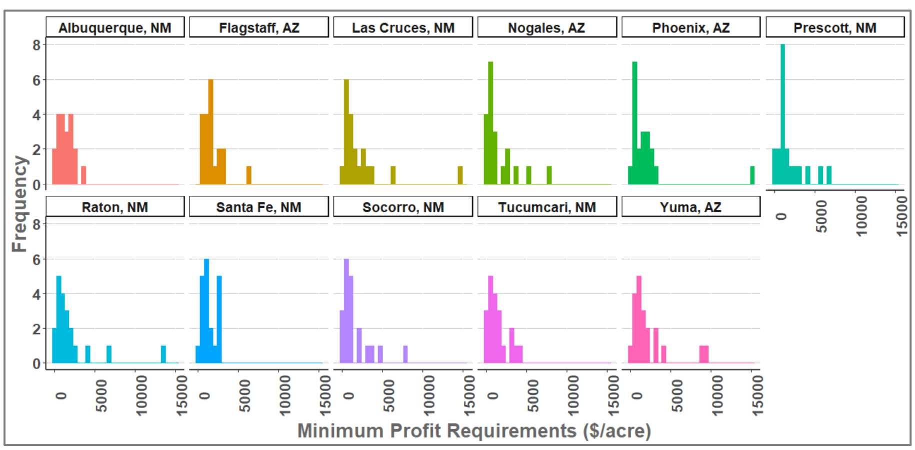


Figure 6-7: Histogram of Lognormally Dist. Minimum Profitability per Acre

The generated minimum profitability levels for twenty different farmers are presented in Figure 6-7. One can observe that within each region, the generated values are dispersed across a range between $0 to $10,000. From this group of farmers, the optimization selects those willing to participate after consideration of the complete set of production and logistic decisions. An expansion of the set of profiles would permit the user to assess the relationship between farmers' profitability profiles and the contract selection.

6.2.3 Results from Case Study

Using information derived from labor and profitability requirements and contract construction, an instance of the optimization is solved. As mentioned earlier, one assumes that there are twenty potential farmers within each of the regions with access to the 10 contract alternatives delineated in Figure 6-6. Those farmers willing to participate would select a single contract from the ten alternatives most profitable to them based on their own minimum profitability value. On the other side, the investor would seek to maximize his/her own profit margins by entering into contracts with a set of farmers within each region that is most profitable to him/her. For this case study, it is assumed that the planting and harvesting schedule is determined by the set of opportunities delineated in section 4.6 and summarized in Figure 6-8. In this case, based on this production schedule the investor would provide shade and greenhouse technologies to selected regions, as well as necessary development for open-field implementations. One also assumes that distribution centers would be installed in Phoenix, Tucson, and Albuquerque, in which the main investor would be responsible for transporting products from each region to the final end consumer. Finally, it is assumed that the investor will be responsible for 10% of the investment cost protected and greenhouse technologies, while the individual farmer would be responsible for the other 90%. In the case of open-field implementation, the investment entity will be responsible for the full cost. Once the different optimization inputs have been considered, including the planting/harvesting schedules, labor and profitability requirements, and contract alternatives, the optimization formulation is solved.

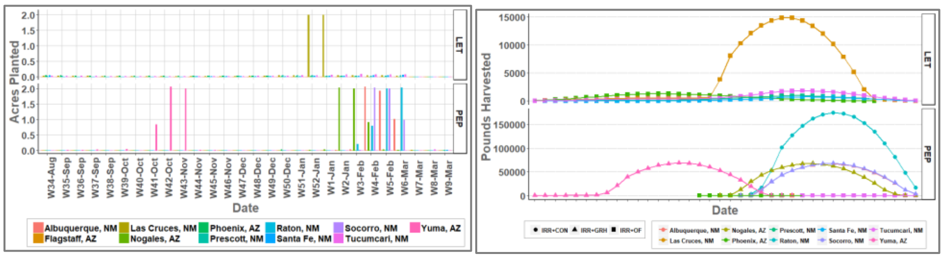


Figure 6-8: Planting/Harvesting Schedule for Regional Farmers

The initial output from the optimization formulation is the contract selection for farmers within each region that decided to participate. This output is presented in Figure 6-9. This figure highlights in black those farmers from Figure 6-7 that decided to participate in production. From this figure one can observe the general profiles of participating farmers within each region. One can observe that the there is a tendency to spread production across multiple farmers within the different regions with lower profitability requirements. In this case, one can use information derived from these farming profiles to design contracts that could improve the general profitability of the region.

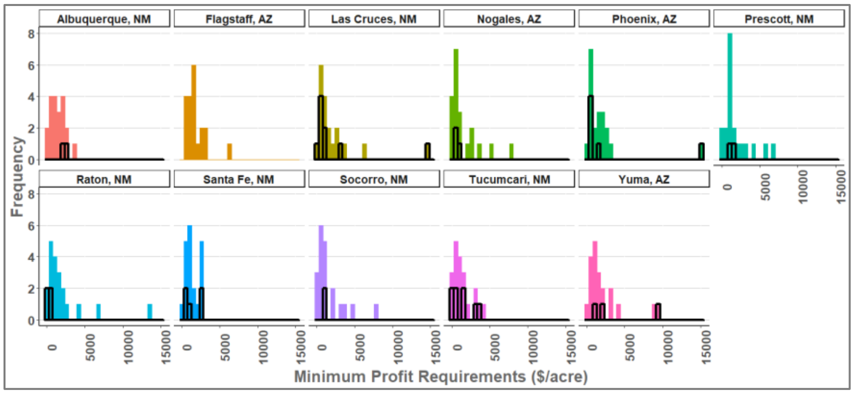


Figure 6-9: Farmer Participation within Each Region

In order to the meet the planting and harvesting schedule presented in Figure 6-8, individual farmers within each region would be tasked with planning labor resource use to meet their contract obligation. Figure 6-10 presents the number of laborers required within each region by type (i.e. full-time, temporary) throughout the season. In this figure, one can observe that labor requirements within each region is mostly dictated by temporary workforce with a single full-time laborer. As to be expected, the number of temporary laborers within each region also increases as the harvesting load is increased.

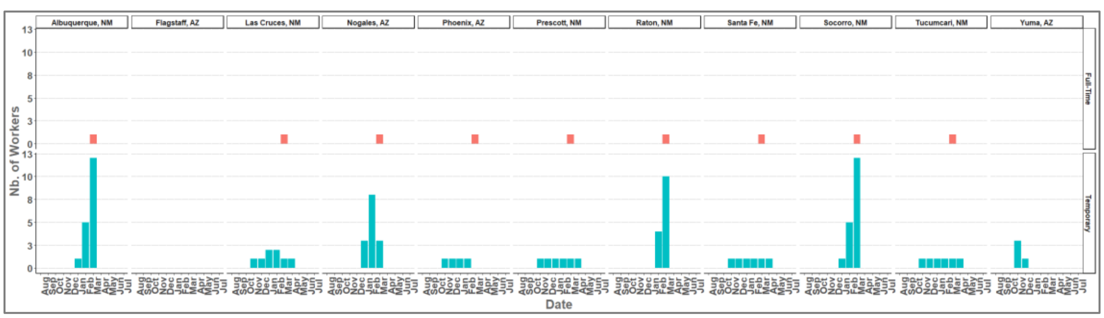


Figure 6-10: Laborers per Region by Type (Full-Time/Temporary)

The aggregated number of laborers per region is presented in Table 6-2. In this case, one can observe that Albuquerque, Nogales, Raton, and Socorro need more than 20 laborers during their production period. Yuma and Las Cruces would require roughly around 3-4 temporary laborers throughout the season. The rest of the regions would require sporadic work from both temporary and full-time laborers. In this aspect, the availability of temporary and full-time laborers within each region may also play a role in their ability to produce higher quantities.

Table 6-2: Number of Laborers by Type

Zone	Full-Time	Temporary
Albuquerque, NM	1	24
Flagstaff, AZ	0	0
Las Cruces, NM	1	4
Nogales, AZ	1	25
Phoenix, AZ	1	1
Prescott, NM	1	1
Raton, NM	1	23
Santa Fe, NM	1	1
Socorro, NM	1	24
Tucumcari, NM	1	1
Yuma, AZ	0	3

The final component is the profitability for each of the participating farmers in the operation, which is presented in Table 6-3. As one can observe from this table, farmers that commit to production are compensated with a higher level of profitability. From this table one can also observe that farmers tend to commit to higher production levels than the one given by the contract, which can become very profitable. Open-field production is distributed among a larger number of farmers. One can also deduce that production under controlled settings would have the highest rewards for participating and is concentrated among fewer farmers. However, it is expected that these individuals would be required to have higher operational experience in managing more sophisticated technologies.

Table 6-3: Profitability per Farmer in each Region

Farmer	Actual	Lettuce/ OF	Pepper/ Control	Pepper/ Protect	Pepper/ OF	Total Acreage
4	Santa Fe, NM	$40	0	0	0	0.1
5	Santa Fe, NM	$75	0	0	0	0.1
12	Santa Fe, NM	$309	0	0	0	0.1
19	Santa Fe, NM	$281	0	0	0	0.1
20	Santa Fe, NM	$374	0	0	0	0.1
24	Yuma, AZ	0	0	$385,741	0	2
25	Yuma, AZ	0	0	$402,346	0	2.1
30	Yuma, AZ	0	0	$151,216	0	0.8
54	Phoenix, AZ	$1,080	0	0	0	0.1
57	Phoenix, AZ	$507	0	0	0	0.1
60	Phoenix, AZ	$551	0	0	$921	0.1
78	Prescott, NM	$916	0	0	0	0.3
80	Prescott, NM	$471	0	0	0	0.1
93	Albuquerque, NM	0	0	$492,875	0	3
95	Albuquerque, NM	0	0	$360,766	0	2
121	Tucumcari, NM	$1,934	0	0	0	0.2
123	Tucumcari, NM	$930	0	0	$611	0.1
125	Tucumcari, NM	$711	0	0	0	0.1
129	Tucumcari, NM	$619	0	0	0	0.1
136	Tucumcari, NM	$345	0	0	0	0.1
137	Tucumcari, NM	$1,789	0	0	0	0.2
139	Tucumcari, NM	$578	0	0	0	0.1
141	Las Cruces, NM	$452	0	0	0	0.1
142	Las Cruces, NM	$28,539	0	0	0	2
149	Las Cruces, NM	$25,148	0	0	$722	2
151	Las Cruces, NM	$1,173	0	0	0	0.1
152	Las Cruces, NM	$648	0	0	0	0.1
154	Las Cruces, NM	$330	0	0	0	0.1
155	Las Cruces, NM	$1,061	0	0	0	0.1
156	Las Cruces, NM	$328	0	0	0	0.1
160	Las Cruces, NM	$745	0	0	0	0.1
165	Nogales, AZ	0	0	$383,141	0	2.1
173	Nogales, AZ	0	0	$323,528	0	1.9
180	Nogales, AZ	0	0	$154,703	0	0.9
192	Raton, NM	0	$230,341	0	0	2.5
194	Raton, NM	0	$30,590	0	0	0.3
195	Raton, NM	0	$20,051	0	0	0.2
198	Raton, NM	0	$179,916	0	0	2
204	Socorro, NM	0	0	$360,192	0	2
206	Socorro, NM	0	0	$503,601	0	3

From this table, one can deduce that regional development of agricultural capabilities can be very profitable if the labor conditions within the different regions are met. It can also be observed that simple investments in protective technologies can be a viable option, and in fact, prove to be more profitable than highly controlled environments. In this case, one can also observe high profit disparities between technologies and regions, which could be addressed with additional constraints in the optimization in order to homogenize overall profitability. Finally, the overall profitability received by the investor (or the optimal solution of the formulation) is approximately $3.7 million for a complete one-year operation.

Finally, this analysis is concluded by an assessment of the running times of the optimization under different sizes of the problem. Again, it is noted that the optimization runs are solved using CPLEX 12.5.0 optimization suite and coded in AMPL on an Intel Core i7-6700, 3.40GHZ computer with 16.0 GB of memory. To assess the impact of the problem size on the optimization formulation, the number of farmers within each of the regions, as well as the number of contracts was used.

Table 6-4: Characterization of Optimization Results

Farmers per Zone	Vars	Cons	MIP Iterns	B&B Nodes	Cuts			Gap %	CPU (s)
					Cover	Flow	Gomory		
1	136,562	41,939	27,899	0	1	2,605	11	0.00	5
2	147,353	50,134	6,823	0	1	3,905	3	0.01	6
3	158,144	58,329	9,206	0	1	4,975	2	0.01	10
4	168,935	66,524	27,311	154	1	5,223	1	0.01	23
5	179,726	74,719	1323670	5,369	1	4,866	2	0.04	299
10	233,681	115,694	11379372	3,481	1	5,454	2	0.02	511
15	287,636	156,669	73388303	37,372	2	6,047	2	0.02	4124

From this table, one can observe that the optimization formulation is able to handle relatively large problem instances. However, as the number of farmers per region surpass 15, the solving scheme begins to have issues in finding optimal solutions. Nevertheless, for purposes of the initial case study of this dissertation, the instance size does not cause major solution issues. However, one of the options to reduce the size of the problem could be to reduce the number of potential decision variables. For example, the selection of transportation mode does not change for larger problem sizes. This means that the risk that an even better solution would be obtained under another transportation mode is very low, and therefore we can eliminate set of decision variables to improve its convergence speed.

6.3 Discussion on Technology Allocation and Labor Resource Planning

This section presented the final component of this dissertation work, which details the technology and resource sharing aspects of local agricultural production development once a set of regions with agricultural production potential have been identified. The main goal in this third aspect was to construct a decentralized perspective to agricultural development within identified regions by considering each farmer and the investor as individual entities with their own profitability requirements. The formulation sought to consider different production and logistic decisions, in similar fashion as in previous chapters, as well as consider labor management decisions for individual farmers within each region. The main literature contribution of this third component is the development of a decentralized formulation applied to the exploration and agricultural development framework. Also, it provides the foundation for further expansion in many different directions, including the design of different contract mechanisms and implementation of more efficient solving mechanisms. Most importantly, this optimization framework provides the analyst with the ability to design a production environment under different assumptions and configurations settings.

There are several advantages to this component of the dissertation. The first advantage is that it allows users to assess different implementation decisions while also considering farmer's individual profitability requirements. This might include the development of different contract designs and assessments of their impact on farmer behavior. Other considerations might be the assessment of different cost and profit sharing mechanisms such as investment, labor, and logistic costs between the main investor and individual farmers of different regions. This would allow the overall profitability of the complementary systems to be shared among all participants. The second advantage to this component is that it sets the primary framework for further decomposition of the problem. In this case, one can use the decentralized formulation to expand the on the number of farmers that can be assessed in order to further profile different types of behavior.

There are several areas of improvement, as well. For example, one of the areas of improvement is in the development of more sophisticated contract designs. This might include the use of risk-based measurements of cost for individual contract selections, in which higher production levels receive weighted rewards when compared to lower commitments. Another area of improvement might be in the cost and profit sharing mechanisms among individual farmers in order to homogenize rewards between participating parties. Lastly the development of more efficient solving mechanisms could easily be applied to this framework given the decomposable structure of the formulations. As it will be mentioned in the next section, these activities are left as part of the future work of this dissertation.

7. DISCUSSION AND FUTURE WORK

There are four main areas of research contribution associated with this dissertation work. The first important component is the development of an overall optimization framework that considers environmental conditions, market prices, plant physiological requirements, and logistic components in the identification of geographical regions with potential to produce high-value agricultural products. This integrated discovery framework is an important advancement in the agricultural planning research as it combines the exploratory yield assessment methodology of fresh vegetables with required supply chain consideration of high-value perishable production. Furthermore, a key advancement in this dissertation is that it limits the dependence of yield estimations of high-value crops on historical information. This deviates from previous works by allowing the construction of planting and harvesting schedules for high-value products based solely on temperature and precipitation patterns by matching physiological requirements to the environmental characteristics of geographical regions. This in turn expands the scope of previous horticultural planning models to a larger geographical area and increases the granularity of yield estimation methodologies previously developed to assess low-value crop potential.

The base optimization framework developed in this dissertation sets the foundation for different areas of expansion. For example, a specific area of improvement is the precision of the yield estimation methodology, which currently depends solely on regional temperature conditions. Additional components to this estimation methodology would be adding the effect of soil properties and sunlight hour availability to yield estimates. Nonetheless, caution must be given to the scarcity of more precise environmental data on a larger scale, which may limit the scope of the assessment. Another area of improvement is in the construction of clusters for temperature and and precipitation patterns within the different regions. With regards to optimization formulation, an interesting area of future work would be the implementation of more efficient solving schemes. It is noted that the optimization formulation has an easily decomposable structure (e.g. independent geographical regions), which can facilitate the implementation of column-generation techniques that would selectively include additional participants (i.e. variables) into the solution space and as a result increase the number of regions that could be assessed simultaneously. However, as before, the availability of additional information needed, such as production and logistic costs for each individual region, may require a larger data collection effort.

The second area of contribution is the incorporation of variability into the initial yield estimation method that is directly derived from environmental parameters. This contribution includes the implementation of stochastic decomposition methods that can handle a larger set of discretized scenarios. As part of this dissertation, two variations of a two-stage stochastic solving schemes were implemented to assess parameter variability. The first approach was the use of L-shaped, deterministic cutting plane methods based on a fixed number of discretized environmental scenarios. However, this approach had various limitations given the size of our particular problem, in which more than one planning region is assessed. The second approach was the use of stochastic, cutting plane methods, which greatly reduces the size of the formulation by eliminating the need to keep full information of previously generated optimality cuts. Also, since the size of the sample considered in the optimization grows alongside each iteration, it allows the ability to assess the individual impact of scenario instances on the solution. Most importantly, this structure is used to incorporate a machine learning component that learns the relationship between first-stage solutions and generated scenarios.

Future work with regards to the stochastic component framework is the refinement of the yield generating method. In this case, an underlying assumption that was made is that random temperature and precipitation samples generated for a given region would be independent from one another. This assumption can be improved by adding a level of correlation over the randomly generated weather samples. Another interesting extension would be the decentralization of the optimization problem, in which each farmer is treated as a unique entity within the actual stochastic optimization set-up. In this case, the size of the problem would have to consider the large number of created scenarios, while also expanding on the number of participants in the optimization. This would most likely require the implementation of more efficient solving schemes to approximate the optimal solution under this particular set-up.

The third main contribution component is the inclusion of machine learning techniques to learn the relationship between first-stage solutions and generated yield scenario instances within a stochastic decomposition implementation. This approach takes advantage of the general solving scheme structure of stochastic decomposition by iteratively training support vector machines model to assess individual scenario instances. In this approach, support vector machine models were used to estimate the probability that each combination selection will enter the first-stage solution in the next iteration based on previous results.

Furthermore, feature reduction techniques are used to facilitate the training of support vector machine models. These models were then used to construct synthetic yield instances that would improve their likelihood of belonging to the first-stage solution. To the best of our knowledge, this dissertation is the first work that incorporates the structure of support vector machines to construct synthetic scenarios within a stochastic decomposition approach.

Future work related to the inclusion of machine learning techniques to a stochastic decomposition framework would be the development of more exact probability estimates of individual scenarios. As noted in this dissertation, the probability estimate was based on previous interactions between yields scenarios and first-stage solutions. This probability estimate is based on the distance between the yield instance in the reduced dimensionality space and the hyperplane separating those solutions that have entered the first-stage solution. Improved methods could incorporate additional aspects specific of the optimization that would provide more insight on these probability estimates. Another improvement on this framework would be the use of machine learning techniques to guide the convergence of the first-stage solution. However, caution must be given to the independency requirements between generated yield instances in a stochastic decomposition set-up, which would have to be addressed in order to maintain its convergence properties.

The fourth component is the development of a decentralized optimization framework that takes the solution from previous problem instances to assign technology resource use and plan labor requirements within each region. This final component considers each farmer to be an individual entity with his/her own minimum profitability requirements and level of participation, and in which the investor seeks to maximize the profitability of his/her investment. This aspect of the dissertation advances previous research by combining individual profitability requirements within an agricultural planning framework. Furthermore, this formulation provides the basic framework through which one can design different contract, labor and cost sharing mechanisms in regional agricultural development. Future work related to the last component of the dissertation would be the development of more efficient solving schemes to consider additional farmer profiles. By analyzing an expanded optimization framework, one can characterize the relationship between the type of operation in each region and different farming profiles. This in turn can be used by investment entities to understand how to construct more targeted contracts to better meet farmers' requirements. This would be especially important if future work considers the risk aspect of horticultural production within the technology assignment formulation.

The overall goal of this dissertation was to develop a set of methods and frameworks aimed at assessing and identifying production potential within different geographical regions. The ability to identify and assess these opportunities can entice micro and small farmers to participate in agricultural supply chains, as well as incentivize external investments into new production implementations. The ultimate purpose is to motivate the connection between investment entities and farmers, in such a way that they can both capitalize on market opportunities in the market place. Also, the idea is for policy makers, farmers, investors, and logistic entities to use this kind of tools to identify potential regions that could be inserted into established distribution and marketing channels by strategically investing in their development. This in turn can have a beneficial impact on marginalized farmers currently excluded from consumption markets lacking the financial ability to improve their supply chain capabilities. We believe that the analytical approach taken to identify and aid agricultural communities could ultimately result in an increase of their production capability, improve regional economies, and decrease rural poverty.

REFERENCES

Adams, S. R., K. E. Cockshull, and C. R. J. Cave. 2001. "Effect of Temperature on the Growth and Development of Tomato Fruits." *Annals of Botany* 88 (5):869–77. https://doi.org/10.1006/anbo.2001.1524.

Ahumada, Omar, J. Rene Villalobos, and A. Nicholas Mason. 2012. "Tactical Planning of the Production and Distribution of Fresh Agricultural Products under Uncertainty." *Agricultural Systems* 112 (October):17–26. https://doi.org/10.1016/j.agsy.2012.06.002.

Ahumada, Omar, and J. Rene Villalobos. 2009. "Application of Planning Models in the Agri-Food Supply Chain: A Review." *European Journal of Operational Research* 196 (1):1–20. https://doi.org/10.1016/j.ejor.2008.02.014.

———. 2011. "Operational Model for Planning the Harvest and Distribution of Perishable Agricultural Products." *International Journal of Production Economics*, Towards High Performance Manufacturing, 133 (2):677–87. https://doi.org/10.1016/j.ijpe.2011.05.015.

Anderson, W. K. 2010. "Closing the Gap between Actual and Potential Yield of Rainfed Wheat. The Impacts of Environment, Management and Cultivar." *Field Crops Research* 116 (1–2):14–22. https://doi.org/10.1016/j.fcr.2009.11.016.

Awudu, Iddrisu, and Jun Zhang. 2012. "Uncertainties and Sustainability Concepts in Biofuel Supply Chain Management: A Review." *Renewable and Sustainable Energy Reviews* 16 (2):1359–68. https://doi.org/10.1016/j.rser.2011.10.016.

Bai, Yun, Yanfeng Ouyang, and Jong-Shi Pang. 2012. "Biofuel Supply Chain Design under Competitive Agricultural Land Use and Feedstock Market Equilibrium." *Energy Economics* 34 (5):1623–33. https://doi.org/10.1016/j.eneco.2012.01.003.

Bailey, T. Glenn, Paul A. Jensen, and David P. Morton. 1999. "Response Surface Analysis of Two-Stage Stochastic Linear Programming with Recourse." *Naval Research Logistics (NRL)* 46 (7):753–76. https://doi.org/10.1002/(SICI)1520-6750(199910)46:7<753::AID-NAV1>3.0.CO;2-M.

Barnett, Douglas, Brian Blake, and Bruce A. McCarl. 1982. "Goal Programming via Multidimensional Scaling Applied to Senegalese Subsistence Farms." *American Journal of Agricultural Economics* 64 (4):720–27. https://doi.org/10.2307/1240581.

Bayraksan, Güzin, and David P. Morton. 2011. “A Sequential Sampling Procedure for Stochastic Programming.” *Operations Research* 59 (4):898–913. https://doi.org/10.1287/opre.1110.0926.

Birge, John, and François Louveaux. 2000. *Introduction to Stochastic Programming*. Springer Science & Business Media.

Birge, John R., and François Louveaux. 1997. *Introduction to Stochastic Programming*. Springer Science & Business Media.

Bishop, Christopher M. 2006. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Secaucus, NJ, USA: Springer-Verlag New York, Inc.

Bocco, Mónica, and Silvina & Tártara Sayago. n.d. “Multicriteria Models: An Application for the Selection of Productive Alternatives.” *Agricultura Técnica* 62 (3):450–62.

Boer, A. John De, and Satish Chandra. 1978. “A Model of Crop Selection in Semi-Subsistence Agriculture and an Application to Mixed Agriculture in Fiji.” *American Journal of Agricultural Economics* 60 (3):436–44. https://doi.org/10.2307/1239940.

Bouman, B. A. M, R. A Schipper, A Nieuwenhuyse, H Hengsdijk, and H. G. P Jansen. 1998. “Quantifying Economic and Biophysical Sustainability Trade-Offs in Land Use Exploration at the Regional Level: A Case Study for the Northern Atlantic Zone of Costa Rica.” *Ecological Modelling* 114 (1):95–109. https://doi.org/10.1016/S0304-3800(98)00121-5.

Brunelli, Ricardo, and Christian von Lücken. 2009. “Optimal Crops Selection Using Multiobjective Evolutionary Algorithms.” *AI Magazine* 30 (2):96. https://doi.org/10.1609/aimag.v30i2.2212.

Carter, Michael R. 1987. “Risk Sharing and Incentives in the Decollectivization of Agriculture.” *Oxford Economic Papers*, New Series, 39 (3):577–95.

Chen, Michael, and Sanjay Mehrotra. 2007. “Epi-Convergent Scenario Generation Method for Stochastic Problems via Sparse Grid.” Technical Report 2007-08. Northwestern University. https://www.researchgate.net/profile/Sanjay_Mehrotra/publication/251793235_Epi-convergent_Scenario_Generation_Method_for_Stochastic_Problems_via_Sparse_Grid/links/02e7e52b4a99e6587e000000.pdf.

Chetty, S., and A.O. Adewumi. 2014. “Comparison Study of Swarm Intelligence Techniques for the Annual Crop Planning Problem.” *IEEE Transactions on*

Evolutionary Computation 18 (2):258–68. https://doi.org/10.1109/TEVC.2013.2256427.

Chung, T. H., and J. C. Spall. 2015. “Integrated Stochastic Optimization and Statistical Experimental Design for Multi-Robot Target Tracking.” In *2015 Winter Simulation Conference (WSC)*, 2463–74. https://doi.org/10.1109/WSC.2015.7408357.

Cook, Roberta L. 2011. “Fundamental Forces Affecting U.S. Fresh Produce Growers And Marketers.” *Choices* 26 (4). http://ideas.repec.org/a/ags/aaeach/120008.html.

Criddle, Richard S., Bruce N. Smith, and Lee D. Hansen. 1997. “A Respiration Based Description of Plant Growth Rate Responses to Temperature.” *Planta* 201 (4):441–45. https://doi.org/10.1007/s004250050087.

Defourny, Boris, Damien Ernst, and Louis Wehenkel. 2012. “Scenario Trees and Policy Selection for Multistage Stochastic Programming Using Machine Learning.” *INFORMS Journal on Computing* 25 (3):488–501. https://doi.org/10.1287/ijoc.1120.0516.

Dietrich, Jan Philipp, Christoph Schmitz, Hermann Lotze-Campen, Alexander Popp, and Christoph Müller. 2014. “Forecasting Technological Change in agriculture—An Endogenous Implementation in a Global Land Use Model.” *Technological Forecasting and Social Change* 81 (January):236–49. https://doi.org/10.1016/j.techfore.2013.02.003.

Dietrich, Jan Philipp, Christoph Schmitz, Christoph Müller, Marianela Fader, Hermann Lotze-Campen, and Alexander Popp. 2012. “Measuring Agricultural Land-Use Intensity – A Global Analysis Using a Model-Assisted Approach.” *Ecological Modelling* 232 (May):109–18. https://doi.org/10.1016/j.ecolmodel.2012.03.002.

Dinar, Ariel, and Robert Mendelsohn. 2011. *Handbook on Climate Change and Agriculture*. Edward Elgar Publishing. http://www.elgaronline.com/view/9781849801164.xml.

Dogliotti, S., M. K. van Ittersum, and W. A. H. Rossing. 2005. “A Method for Exploring Sustainable Development Options at Farm Scale: A Case Study for Vegetable Farms in South Uruguay.” *Agricultural Systems* 86 (1):29–51. https://doi.org/10.1016/j.agsy.2004.08.002.

Drost, Dan. 2010. “Lettuce in the Garden.” Home Gardening. Utah State University: Cooperative Extension.

Federgruen, Awi, Upmanu Lall, and Simsek Serdar. 2014. “Supply Chain Analysis of Contract Farming.”

Flores, Hector, and J. Rene Villalobos. 2013. “Using Market Intelligence for the Opportunistic Shipping of Fresh Produce.” *International Journal of Production Economics* 142 (1):89–97.

Gibbons, J. M., and S. J. Ramsden. 2008. “Integrated Modelling of Farm Adaptation to Climate Change in East Anglia, UK: Scaling and Farmer Decision Making.” *Agriculture, Ecosystems & Environment* 127 (1–2):126–34. https://doi.org/10.1016/j.agee.2008.03.010.

Gjerdrum, Jonatan, Nilay Shah, and Lazaros G. Papageorgiou. 2001. “Transfer Prices for Multienterprise Supply Chain Optimization.” *Industrial & Engineering Chemistry Research* 40 (7):1650–60. https://doi.org/10.1021/ie000668m.

Glen, John J. 1987. “Mathematical Models in Farm Planning: A Survey.” *Oper. Res.* 35 (5):641–666. https://doi.org/10.1287/opre.35.5.641.

Harrison, Kerry. 2014. “Commercial Tomato Production Handbook.” B 1312. Irrigation. UGA Extension: University of Georgia. http://extension.uga.edu/publications/detail.cfm?number=B1312.

Hartz, Tim, Marita Cantwell, Michelle LeStrange, Richard Smith, Jose Aguiar, and Oleg Daugovish. 2008. “Bell Pepper Production in California.” Publication 7217. Vegetable Production Series. UC Vegetable Research & Information Center: University of California: Division of Agriculture and Natural Resources.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. 2009. *Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd Edition.* Second Edition. Springer Series in Statistics. http://statweb.stanford.edu/~tibs/ElemStatLearn/.

Higle, Julia L., and Suvrajeet Sen. 1996. *Stochastic Decomposition: A Statistical Method for Large Scale Stochastic Linear Programming*. Springer Science & Business Media.

Hoang, Viet-Ngu. 2013. “Analysis of Productive Performance of Crop Production Systems: An Integrated Analytical Framework.” *Agricultural Systems* 116 (March):16–24. https://doi.org/10.1016/j.agsy.2012.12.005.

Huh, Woonghee Tim, and Upmanu Lall. 2013. “Optimal Crop Choice, Irrigation Allocation, and the Impact of Contract Farming.” *Production and Operations Management* 22 (5):1126–43. https://doi.org/10.1111/poms.12007.

Ittersum, Martin K. van, Kenneth G. Cassman, Patricio Grassini, Joost Wolf, Pablo Tittonell, and Zvi Hochman. 2013. “Yield Gap Analysis with Local to Global relevance—A Review.” *Field Crops Research*, Crop Yield Gap Analysis – Rationale, Methods and Applications, 143 (March):4–17. https://doi.org/10.1016/j.fcr.2012.09.009.

Janssen, Sander, and Martin K. van Ittersum. 2007. “Assessing Farm Innovations and Responses to Policies: A Review of Bio-Economic Farm Models.” *Agricultural Systems*, Special Section: sustainable resource management and policy options for rice ecosystems International symposium on sustainable resource management and policy options for rice ecosystems, 94 (3):622–36. https://doi.org/10.1016/j.agsy.2007.03.001.

Johnson, Richard A., and Dean W. Wichern. 1992. *Applied Multivariate Statistical Analysis*. 3rd edition. Englewood Cliffs, NJ: Prentice Hall.

Johnson, Scott L., Richard M. Adams, and Gregory M. Perry. 1991. “The On-Farm Costs of Reducing Groundwater Pollution.” *American Journal of Agricultural Economics* 73 (4):1063–73. https://doi.org/10.2307/1242434.

Just, Richard E., and David Zilberman. 1983. “Stochastic Structure, Farm Size and Technology Adoption in Developing Agriculture.” *Oxford Economic Papers*, New Series, 35 (2):307–28.

Kaiser, Cheryl, and Matt Ernst. 2014. “Bell Peppers.” Cooperative Extension Service: University of Kentucky College of Agriculture. https://www.uky.edu/Ag/CCD/introsheets/bellpeppers.pdf.

Key, Nigel, and David Runsten. 1999. “Contract Farming, Smallholders, and Rural Development in Latin America: The Organization of Agroprocessing Firms and the Scale of Outgrower Production.” *World Development* 27 (2):381–401. https://doi.org/10.1016/S0305-750X(98)00144-2.

Kleywegt, A., A. Shapiro, and T. Homem-de-Mello. 2002. “The Sample Average Approximation Method for Stochastic Discrete Optimization.” *SIAM Journal on Optimization* 12 (2):479–502. https://doi.org/10.1137/S1052623499363220.

Law, Averill M., and W. David Kelton. 1991. *Simulation Modeling and Analysis*. McGraw-Hill.

LeBoeuf, Janice. 2004. “The Effect of Extreme Temperatures on the Tomato and Pepper Crop.” Ontario, Canada: Ministry of Agriculture, Food, and Rural Affairs. http://www.omafra.gov.on.ca/english/crops/facts/info_tomtemp.htm.

Lotze-Campen, Hermann, Christoph Müller, Alberte Bondeau, Stefanie Rost, Alexander Popp, and Wolfgang Lucht. 2008. “Global Food Demand, Productivity Growth, and the Scarcity of Land and Water Resources: A Spatially Explicit Mathematical Programming Approach.” *Agricultural Economics* 39 (3):325–38. https://doi.org/10.1111/j.1574-0862.2008.00336.x.

Lowe, Timothy J., and Paul V. Preckel. 2004. “Decision Technologies for Agribusiness Problems: A Brief Review of Selected Literature and a Call for Research.” *Manufacturing & Service Operations Management* 6 (3):201–8. https://doi.org/10.1287/msom.1040.0051.

Lu, C. H., M. K. van Ittersum, and R. Rabbinge. 2004. “A Scenario Exploration of Strategic Land Use Options for the Loess Plateau in Northern China.” *Agricultural Systems* 79 (2):145–70. https://doi.org/10.1016/S0308-521X(03)00069-6.

Lucas, M. T., and D. Chhajed. 2004. “Applications of Location Analysis in Agriculture: A Survey.” *Journal of the Operational Research Society* 55 (6):561–78. https://doi.org/10.1057/palgrave.jors.2601731.

Markelova, Helen, Ruth Meinzen-Dick, Jon Hellin, and Stephan Dohrn. 2009. “Collective Action for Smallholder Market Access.” *Food Policy*, Collective Action for Smallholder Market Access, 34 (1):1–7. https://doi.org/10.1016/j.foodpol.2008.10.001.

Masabni, Joe. 2016. “Bell Peppers Crop Guide.” Small Acreage Horticultural Crops. Agrilife Extension: Texas A&M. http://aggie-horticulture.tamu.edu/smallacreage/crops-guides/vegtables/bell-peppers/.

Mason, A.N., and J.R. Villalobos. 2014. “Coordination of Perishable Crop Production Using Auctions Mechanisms.” *Agricultural Systems* Under Review.

Mattson, Neil. 2015. “Controlled Environment Agriculture for Year-round Vegetables: Production Systems, Costs, and Potential Crop Yield.” School of Integrative Plant Science: Cornell University. http://www.hort.cornell.edu/expo/proceedings/2017/Greenhouse%20Vegetables.Controlled%20Environment.pdf.

Minot, Nicholas. 2007. “Contract Farming in Developing Countries: Patterns, Impact, and Policy Implications.” Case Study #6-3. Cornell University.

Müller, Christoph, and Richard D. Robertson. 2014. “Projecting Future Crop Productivity for Global Economic Modeling.” *Agricultural Economics* 45 (1):37–50. https://doi.org/10.1111/agec.12088.

Nagarajan, Mahesh, and Greys Sošić. 2008. “Game-Theoretic Analysis of Cooperation among Supply Chain Agents: Review and Extensions.” *European Journal of Operational Research* 187 (3):719–45. https://doi.org/10.1016/j.ejor.2006.05.045.

Neumann, Kathleen, Peter H. Verburg, Elke Stehfest, and Christoph Müller. 2010. “The Yield Gap of Global Grain Production: A Spatial Analysis.” *Agricultural Systems* 103 (5):316–26. https://doi.org/10.1016/j.agsy.2010.02.004.

NMED. 2016. “New Mexico: Water and Sewer Rate Surveys.” NMED Drinking Water Bureau: New Mexico Environment Department. https://www.env.nm.gov/dwb/rate.htm.

NOAA. 2016a. “Climate Data Online: Dataset Discovery.” National Centers for Environmental Information: National Oceanic and Atmospheric Administration.

———. 2016b. “Land-Based Station Data.” Climate Data Online. National Centers for Environmental Information: National Oceanic and Atmospheric Administration. https://www.ncdc.noaa.gov/data-access/land-based-station-data.

Orzolek, Michael, Steven M. Bogash, Matthew Harsh, and Jayson Harper. 2016. “Tomato Production.” Ag. Alternatives. PennState Extension: Penn State University.

Pal, Bijay Baran, Debjani Chakraborti, and Papun Biswas. 2009. “A Genetic Algorithm Based Hybrid Goal Programming Approach to Land Allocation Problem for Optimal Cropping Plan in Agricultural System.” In , 181–86. IEEE. https://doi.org/10.1109/ICIINFS.2009.5429867.

Platt, John C. 1999. “Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods.” In *Advances in Large Margin Classifiers*, 61–74. MIT Press.

Santoso, Tjendera, Shabbir Ahmed, Marc Goetschalckx, and Alexander Shapiro. 2005. “A Stochastic Programming Approach for Supply Chain Network Design under Uncertainty.” *European Journal of Operational Research* 167 (1):96–115. https://doi.org/10.1016/j.ejor.2004.01.046.

Schmitz, Christoph, Anne Biewald, Hermann Lotze-Campen, Alexander Popp, Jan Philipp Dietrich, Benjamin Bodirsky, Michael Krause, and Isabelle Weindl. 2012. “Trading More Food: Implications for Land Use, Greenhouse Gas Emissions, and the Food System.” *Global Environmental Change* 22 (1):189–209. https://doi.org/10.1016/j.gloenvcha.2011.09.013.

Smith, Powell, Bob Polomski, and Debbie Shaughnessy. 2003. “Lettuce.” HGIC 1312. Clemson: Cooperative Extension.

Stoddard, Scott, Michelle LeStrange, Brenna Aegerter, Karen Klonsky, and Richard De Moura. 2007. “Sample Costs to Produce Fresh Market Tomatoes.” TM-SJ-07. San Joaquin Valley: University of California: UC Cooperative Extension.

Tan, Barış, and Nihan Çömden. 2012. “Agricultural Planning of Annual Plants under Demand, Maturation, Harvest, and Yield Risk.” *European Journal of Operational Research* 220 (2):539–49. https://doi.org/10.1016/j.ejor.2012.02.005.

Tao, Fulu, Masayuki Yokozawa, and Zhao Zhang. 2009. “Modelling the Impacts of Weather and Climate Variability on Crop Productivity over a Large Area: A New Process-Based Model Development, Optimization, and Uncertainties Analysis.” *Agricultural and Forest Meteorology* 149 (5):831–50. https://doi.org/10.1016/j.agrformet.2008.11.004.

Tewari, J.P. 2015. “Guide to Commercial Greenhouse Sweet Bell Pepper Production in Alberta.” Alberta: Agriculture and Forestry: University of Alberta.

Tourte, Laura, Richard Smith, Karen Klonsky, Dan Sumner, Christine Gutierrez, and Don Stewart. 2015. “Sample Costs to Produce and Harvest Romaine Hearts.” Central Coast Region. Agriculture and Natural Resources – Agricultural Issues Center: University of California Cooperative Extension. https://coststudyfiles.ucdavis.edu/uploads/cs_public/53/a0/53a09cf4-9732-451e-b5ad-1beb770adbd0/2015romainehearts-finaldraft_1-27-2016.pdf.

USDA. 2012. “2012 Census of Agriculture County Profile: Yuma County.” U.S. Census of Agriculture: United States Department of Agriculture: US Department of Agriculture. http://www.agcensus.usda.gov/Publications/2012/Full_Report/Volume_1,_Chapter_1_State_Level/Arizona/st04_1_009_010.pdf.

———. 2013. “U.S. per Capita Consumption of Fresh Vegetables in the United States in 2013, by Vegetable Type (in Pounds).” Yearbook Tables. Economic Research Service: US Department of Agriculture.

———. 2015a. “Lettuce.” Agricultural Marketing Resource Center: USDA Rural Development. http://www.agmrc.org/commodities-products/vegetables/lettuce/.

———. 2015b. “Fruit and Vegetable Prices.” Economic Research Service: US Department of Agriculture. http://www.ers.usda.gov/data-products/fruit-and-vegetable-prices.aspx.

———. 2016. “Fruit and Vegetable Market News: Specialty Crops.” USDA Market News. Agricultural Marketing Services: United States Department of Agriculture. http://www.marketnews.usda.gov.

Vanthoor, B. H. E., P. H. B. de Visser, C. Stanghellini, and E. J. van Henten. 2011. “A Methodology for Model-Based Greenhouse Design: Part 2, Description and Validation of a Tomato Yield Model.” *Biosystems Engineering* 110 (4):378–95. https://doi.org/10.1016/j.biosystemseng.2011.08.005.

Verburg, Peter H, Youqi Chen, and Tom (A.) Veldkamp. 2000. “Spatial Explorations of Land Use Change and Grain Production in China.” *Agriculture, Ecosystems & Environment* 82 (1–3):333–54. https://doi.org/10.1016/S0167-8809(00)00236-X.

Waha, K., C. Müller, A. Bondeau, J. P. Dietrich, P. Kurukulasuriya, J. Heinke, and H. Lotze-Campen. 2013. “Adaptation to Climate Change through the Choice of Cropping System and Sowing Date in Sub-Saharan Africa.” *Global Environmental Change* 23 (1):130–43. https://doi.org/10.1016/j.gloenvcha.2012.11.001.

Wallace, Michael T., and Joan E. Moss. 2002. “Farmer Decision-Making with Conflicting Goals: A Recursive Strategic Programming Analysis.” *Journal of Agricultural Economics* 53 (1):82–100. https://doi.org/10.1111/j.1477-9552.2002.tb00007.x.

Wang, K. -J., S. -M. Wang, and J. -C. Chen. 2008. “A Resource Portfolio Planning Model Using Sampling-Based Stochastic Programming and Genetic Algorithm.” *European Journal of Operational Research* 184 (1):327–40. https://doi.org/10.1016/j.ejor.2006.10.037.

Wart, Justin van, K. Christian Kersebaum, Shaobing Peng, Maribeth Milner, and Kenneth G. Cassman. 2013. “Estimating Crop Yield Potential at Regional to National Scales.” *Field Crops Research*, Crop Yield Gap Analysis – Rationale, Methods and Applications, 143 (March):34–43. https://doi.org/10.1016/j.fcr.2012.11.018.

WIFA. 2015. “Arizona Water and Wastewater Rates and Rate Structures.” Water Infrastructure Finance Authority of Arizona: UNC Environmental Finance Center.

Will, Margaret. 2013. *Contract Farming Handbook: A Practical Guide for Linking Small-Scale Producers and Buyers through Business Model Innovation*. German Federal Ministry for Economic Cooperation and Development. Germany: Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH.

Wishon, C., J. R. Villalobos, N. Mason, H. Flores, and G. Lujan. 2015. “Use of MIP for Planning Temporary Immigrant Farm Labor Force.” *International Journal of*

Production Economics 170, Part A (December):25–33. https://doi.org/10.1016/j.ijpe.2015.09.004.

Wit, C. T. de, H. van Keulen, N. G. Seligman, and I. Spharim. 1988. “Application of Interactive Multiple Goal Programming Techniques for Analysis and Planning of Regional Agricultural Development.” *Agricultural Systems* 26 (3):211–30. https://doi.org/10.1016/0308-521X(88)90012-1.

Yue, Dajun, and Fengqi You. 2014. “Fair Profit Allocation in Supply Chain Optimization with Transfer Price and Revenue Sharing: MINLP Model and Algorithm for Cellulosic Biofuel Supply Chains.” *AIChE Journal* 60 (9):3211–29. https://doi.org/10.1002/aic.14511.

Zander, P, and H Kächele. 1999. “Modelling Multiple Objectives of Land Use for Sustainable Development.” *Agricultural Systems* 59 (3):311–25. https://doi.org/10.1016/S0308-521X(99)00017-7.

Zhang, John. 2016. “Production of Sweet Bell Peppers.” Production Guides. Alberta.ca: Alberta: Agriculture and Forestry. http://www1.agric.gov.ab.ca/$department/deptdocs.nsf/all/opp4523.

Zijun Wang, David J. Leatham, and Thanapat Chaisantikulawat. 2002. “External Equity in Agriculture: Risk Sharing and Incentives in a Principal-agent Relationship.” *Agricultural Finance Review* 62 (1):13–24. https://doi.org/10.1108/00214850280001126.

APPENDIX A

YIELD SCENARIO CHARACTERIZATION

Support vector machines were used to learn and fit a model that attempts to find a relationship between generated yield instances of different production combination alternatives (i.e. crop, region, and technology) and their selection in the first-stage problem. The inclusion of learning techniques into the stochastic decomposition frameworks, along with a feature reduction component, serves to add another level of transparency to its implementation. The decision to use of support vector machine models was due to the representation of yield scenarios in this dissertation. To better explain the series of steps in the application of support vector machines to learn this relationship, a snapshot of a single iteration stochastic decomposition is used. Furthermore, this appendix shows how the structure of support vector models were used to construct synthetic yield scenarios.

Figure I-1 presents the 140th iteration of the stochastic decomposition algorithm. Each panel is a scatter plot of the first two principal components of yield scenarios per crop and technology combination. In this figure, focus is given to open-field, bell pepper production shown in the right-mid panel of the figure, in which each point is a two-principal component representation of previously generated yield instances for a single region. In other words, each point captures compressed information of estimated yields between any two planting weeks for a single scenario in each region. Since this is the 140th iteration, each region has multiple yield scenario points plotted within each crop-technology panel. Additionally, the current 140th iteration for all regions are highlighted with a dark point. Finally, those region, crop, and technology combinations that were part of the 139th first-stage solution are fully labeled with their name. This allows one to identify those combinations that are currently part of the first-stage optimal solution. For example, from this figure, one can observe that only Albuquerque and Tucumcari, NM, have been selected for pepper production under open-field conditions during the 139th iteration. The question is then how to use additional information that can be derived from this set-up so that one can determine the general characteristics of points previously selected by the first-stage problem based on their location in this principal component space.

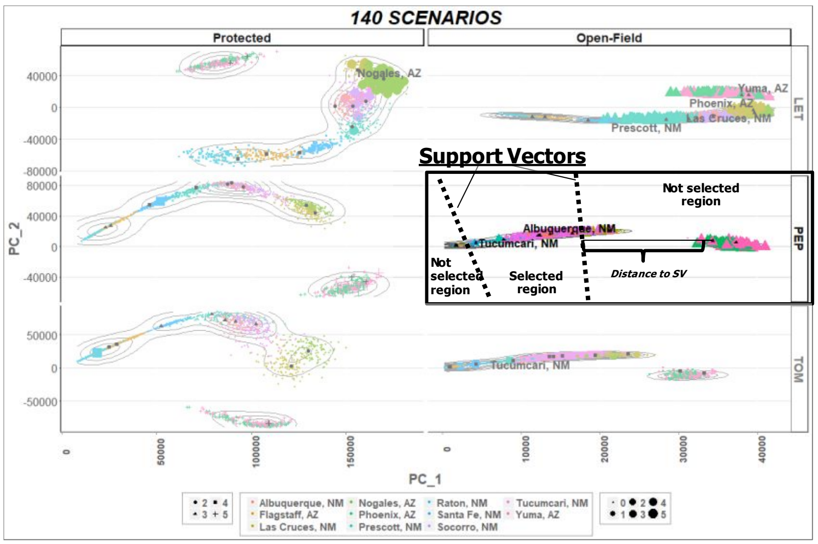


Figure I-1: Application of Support Vectors to Yield Scenario Space

The main task behind the construction of support vector machine (SVM) models is the identification of optimal separating hyperplanes, which serve as boundaries to classify between two main classification classes. For purposes of our own problem, the direct application would be the identification of hyperplanes, or support vectors, that are able to separate between those combinations that were selected in the first-stage solution and those that were not. These support vectors are represented by the dummy dotted lines in Figure I-1, in which the idea would be to determine the characteristics of regions most likely to be selected in the next iteration's first-stage solution. Furthermore, one would want to know how far away each individual combination is from actually being part of the selection region.

In general, the objective of support vector machines is to find separating hyperplanes that separate between two particular classes and maximize the distance between the closest points from either class (Vapnik, 2013). Using the set of definitions and properties highlighted by Hastie et al. (2009), this hyperplane can be represented by the equation $f(x) = \beta_0 + \beta^T x = 0$ and can be summarized by three basic properties. The first basic property is that for any two points x_1 and x_2 on this hyperplane, $\beta^T(x_1 - x_2) = 0$, which means that $\beta/\left||\beta|\right|$ is the vector normal to the hyperplane. The second basic property is that for any x_0 on the hyperplane, $\beta^T x_0 = -\beta_0$. Hence, the signed distance of any point x to the hyperplane can be expressed by $1/\left||\beta|\right|(\beta^T x + \beta_0)$, which is the third property. Therefore, in general $f(x)$ is proportional to the signed distance from x to the hyperplane defined by $f(x) = 0$. Furthermore, the classification (or sign of $f(x)$) is given by $\cos(\theta) = \frac{\dot{\beta}}{||\beta||}(x - x_0)$, where θ is the angle between $\frac{\dot{\beta}}{||\beta||}$ and $(x - x_0)$ and x_0 is any point on the hyperplane. The technical objective is then to solve a non-linear optimization model that finds the hyperplane that maximizes the separation between the two classes based on these basic properties. For a more in-depth explanation of how these basic properties are used to find optimal hyperplanes, the reader is referred to Hastie et al. (2009). An interesting characteristic of SVMs is that one can use the distance between any point to the separating hyperplane as an input to estimate its probability of actually belonging to a particular binary classification. In this case, one of the most widely used technique is one presented by Platt (1999), who provides a post-classification technique based on the fitting of sigmoid functions of decision values outputted by SVM binary classification. The probabilities are estimated by determining the parameters that minimize the negative log-likelihood of the sigmoid function.

Back to the problem at-hand, the application of support vector machine models allows one to split the principal component space of the yield scenarios between those that were selected in the optimal first-stage solution, and those that were not, through the use of optimal hyperplanes. In this case, it may also provide the user with the ability assess the stability of any one particular combination in the first-stage solution. For example, if a combination selection is continuously changing within previous first-stage optimization results, then this would mean, in part, that this combination is closer to the decision boundary vector. On the other hand, if a combination is never selected (or is always selected), then this would mean that the combination is farther away from the decision boundary and its likelihood of changing class during the next iteration's solution is very low. By using the techniques developed by Platt (1999) in an R svm implementation package from Karatzoglou et al. (2006), one can then estimate each combination's probability of belonging to the first-stage solution based on their current distance from the separating hyperplane.

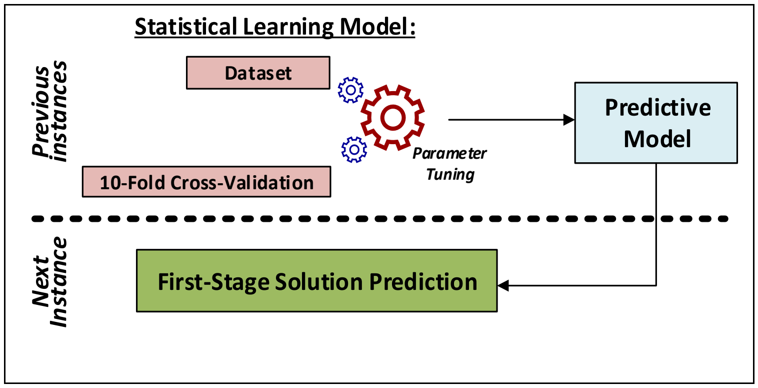


Figure I-2: Statistical Learning Model Training Framework

To perform this set of assessments, a series of intermediate steps are performed to train the support vector model. From Figure I-2, one can observe that a 10-fold cross validation is used to iteratively train the support vector machine model at the end each of each second-stage optimization. The training set for the support vector models are comprised of previously generated yield scenario instances for all combination selections. Once the support vector model is trained, a prediction is given for all combination selections for the current instance. The prediction from the SVM output is then compared against the actual first-stage solution in the upcoming first-stage results. This comparison can then be used as a proxy to gauge the stability of each individual combination.

$$Crop + Tech + Region + PC_1 + PC_2 + PC_3 \sim \mathrm{I}\,(Part\ of\ First\ Stage\ Solution)$$

The support vector model is trained using the principal component representation of the yield scenarios for each crop, technology, and region combination. The output of the model is a binary classification on whether this particular combination was included in the optimal first-stage solution.

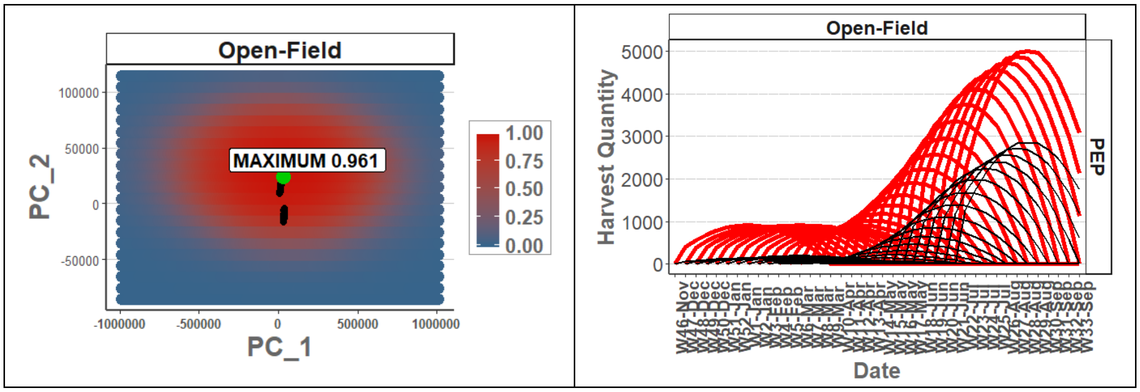


Figure I-3: Use of SVM Models to Estimate Maximum Probability Points

Once an SVM model has been trained, the final component is to attempt to understand relationships between the location of yield scenarios and the probability of being part of the first-stage solution. Thus, the method used was to simply iterate along different principal component values of each selection combination and to use the SVM model to predict their probability at different points. An example of this iteration is presented in the left-hand side of Figure I-3, which captures SVM probability outputs at different points in the principal component space of open-field production for a given region. In this figure, the green point represents the principal component location with the highest probability. The black points represent actual previous yield scenario points. Once this maximum point has been identified, then one can use the principal component scores to revert the compressed information onto an uncompressed form. The uncompressed form of this maximum probability point is given by the red yield pattern on the right-hand side of Figure I-3. The black yield curves are the actual output of the latest generated instance. From this figure, one can determine the type of yield pattern needed in this region in order to improve its changes of entering the first-stage solution.

APPENDIX B

DETERMINISTIC FORMULATION

```
# Formulation

# Data sets #

set WEEKP ordered;       # weeks of the planning period

set WEEKH ordered;       # weeks of the harvesting period

set CROP;       # Crops for planting

set CUST;       # Customers

set TECH;                                               # Growing technologies

set ZON;                   # Zones to be considered

set DC;                                                   # Distribution centers to be considered

set LOC;       # Locations available for planting crops

set MOD;       #  Transportation mode

#####

set WEEK ordered;

set WEEKS ordered;

set WEEK2 within {WEEKP, WEEKH,CROP,ZON,TECH};

set WEEK3 within {WEEKP, WEEKH,CROP,ZON};

set WEEK4 within {WEEKP, WEEKH,CROP};

set WEEK5 within {WEEKP, WEEKH,ZON};

set WEEK6 within {WEEKP, WEEKH};

set WEEK1 within {WEEKH, WEEKS};

set QUAL:= 2..2;                              # Color characteristic of products

########
```

```
# Parameter definition #

# Production/Yield-Related

param YDist {WEEKP,WEEKH,CROP,ZON,TECH} >= 0 # Expected YDist per crop

param Yield {WEEKP,CROP,ZON,TECH} >=0; # Yield production expected

param Salv {WEEKP,WEEKH,CROP,ZON,TECH} >=0;    # Expected waste of crops

param MaxDem {WEEKS,CROP,CUST} >= 0;    #Maximum demand by customer

param MinDem {WEEKS,CROP,CUST} >= 0;    #Minimum demand from customers

param Qmin {CUST}>=0;                              # Quality demanded by customer

param COL {WEEKH,CROP,QUAL}>=0;                     # Binary indicator parameter

param minl{CROP} >=0;                              # Minimum amount to plant per crop

param maxl{CROP} >=0;                              # Maximum amount to plant per crop

param Ctech{TECH, ZON} >=0;              # Cost of technology u for location l

param Coper{TECH, ZON} >= 0;  # Cost per acre for operating technology u in location l

param Cplant{CROP} >=0;          # Cost to plant per crop

param Cwater{ZON}  >=0;                            # Cost of water per location l

param cidloc {LOC} symbolic;

param cidzone{ZON} symbolic;

# Environment-Related

param MinWReq {CROP,TECH} >=0; # Wtr reqrd per acre of crop j using technology u

param LAvail {LOC} >=0;       # Available hectares for planting at location l

param LRainRec{WEEKP,WEEKH,ZON} >= 0;              # Gallons of water received

# Logistics Related
```

```
param SL {CROP} >=0; # Shelf life of product k

param LT {CUST} >=0;# Lead time required by the customer

param Weight {CROP} >=0;  # Quantity in required of crop j to form a case of product k

param PZcap {ZON} >=0;# Capacity of the packaging facility for a time period

param TimeZC {ZON,CUST,MOD} >=0;     #Transportation from zone z to customer c

param TimeDC {DC,CUST,MOD} >=0;          # Time of transportation from DC d to customer i

param TimeZD {ZON,DC} >=0;              # Time of transportation from zone z to DC d

param CTZC{ZON,CUST,MOD} >=0; # Transportation cost from zone z to customer c

param CTZD{ZON,DC} >=0;# Transportation cost from zone z to customer c

param CTDC{DC,CUST,MOD} >=0;     # Transportation cost from zone z to customer c

param CTLZ{ZON} >=0;

param Cw {ZON} >= 0; # Cost of warehouse

param Cd {DC} >= 0; # Cost of DC

param Ccase{CROP} >=0;                          # Cost for packing case for crop j

param TraF{CROP}  >=0;

param WZ_Cap{ZON} >= 0;

# Market Related

param price {WEEKS,CROP,CUST} >= 0;# price for customer i per week t from crop j

# Additional parameters

### Marginal profitability estimates

param Revenue_Profit{CROP,DC,CUST} ;
```

```
param Zone_Plant_Cost {ZON,CROP} >= 0;

param Zone_Water_Cost {ZON,CROP} >= 0;

param Zone_Invst_Cost {ZON,CROP} >= 0;

param Logistic_Cost {CROP} ;

param Tech_Selection {ZON,CROP,TECH} >= 0;

param AvailCap >=0;                    # Capital available for technology investments

########## Variable Definition

#Planting and resource decisions

var Plant {WEEKP,CROP,LOC,TECH} >=0; # Area plant crop j in period p at location l

var MicroHarv {h in WEEKH, j in CROP,  l in LOC ,u in TECH} >=0;    # Harvest (pnds)

var ZoneHarv {(p,h,j,z) in WEEK3}  >=0;           # Harvest (pounds) of crop j

var WatAll {WEEKP,WEEKH,ZON,CROP} >=0;        # Water quantity allocated to zone

var B {CROP,LOC,TECH} >= 0 binary;              # Binary decision to plant crop

#Logistic related variables

var PACK {WEEKH,CROP,QUAL,ZON} >=0;             # Quantity of  crop packed

var SLZ {WEEK1,CROP,QUAL,l in LOC,z in ZON: cidloc[l]==cidzone[z]} >=0;

# Quantitiy to ship from farmer directly to customer i

var SZD {WEEK1,CROP,QUAL,ZON,DC} >=0;

var SZC {WEEK1,CROP,QUAL,ZON,CUST,MOD} >=0;       # Quantity to ship from

#zone's warehouse to customer i

var SDC {WEEK1,CROP,QUAL,DC,CUST,MOD} >= 0;

var Invw {WEEK1,CROP,QUAL,ZON} >=0;  # Inventory in the warehouse
```

```
#within each zone
var Invd {WEEK1,CROP,QUAL,DC} >=0;  # Inventory at the DC
var Add_War{ZON} >=0;
####################
# Objective Function #
maximize Yield_revenue:
# Market revenues
sum {(h,t) in WEEK1, j in CROP, q in QUAL,i in CUST,d in DC, m in MOD}
SDC[h,t,j,q,d,i,m]*price[t,j,i] ##Selling to customer directly from packing facility
 # Planting/production costs
-sum{p in WEEKP,j in CROP,l in LOC,u in TECH} (Plant[p,j,l,u]*Cplant[j])
# Planting Costs (at this point assuming that it does not vary based on technology
-sum {p in WEEKP, z in ZON, u in TECH, j in CROP, l in LOC: cidloc[l]==cidzone[z]}
(Coper[u,z]+Ctech[u,z])*Plant[p,j,l,u]   # Becomes the technology costs
-sum{z in ZON,h in WEEKH,j in CROP, q in QUAL} (PACK[h,j,q,z]/25*Ccase[j])
Packing costs*10
-sum{(h,t) in WEEK1, j in CROP, q in QUAL, z in ZON} Invw[h,t,j,q,z]*Cw[z]
-sum{(h,t) in WEEK1, j in CROP, q in QUAL, d in DC} Invd[h,t,j,q,d]*Cd[d]
-sum{(p,h,z) in WEEK5,j in CROP} (WatAll[p,h,z,j]*Cwater[z])
# Transportation costs
-sum {(h,t) in WEEK1, j in CROP,q in QUAL,d in DC, i in CUST,m in MOD}
SDC[h,t,j,q,d,i,m]*CTDC[d,i,m]          # From facility to customer
```

```
-sum {(h,t) in WEEK1, j in CROP,q in QUAL,z in ZON,d in DC}
SZD[h,t,j,q,z,d]*CTZD[z,d]          # From facility to customer
-sum {h in WEEKH, j in CROP,q in QUAL,l in LOC, z in ZON:cidloc[l]==cidzone[z]}
SLZ[h,h,j,q,l,z]*CTLZ[z]
 #### Keep feasibility
-sum{z in ZON} M*Add_War[z];
#######
######################
##CONSTRAINTS
#
# Planting decisions
subject to TechInvL:                # Investment in Technologies is limited by available capital
 sum{z in ZON, j in CROP, l in LOC,  u in TECH: cidloc[l]== cidzone[z]}
(B[j,l,u]*Ctech[u,z]) <= AvailCap;
subject to Tot_land_Loc {l in LOC,j in CROP,u in TECH}:
sum{p in WEEKP} Plant[p,j,l,u] <= LAvail[l]* B[j,l,u]; #sum{j in CROP, u in TECH} *
B[j,l,u];
subject to Tot_land{l in LOC}:
sum{p in WEEKP,j in CROP, u in TECH} Plant[p,j,l,u] <= LAvail[l];
subject to TechType{j in CROP }:
sum{u in TECH,l in LOC} B[j,l,u] <=5;
subject to M_Prod {p in WEEKP, j in CROP, l in LOC,u in TECH}:
```

```
Plant[p,j,l,u] <=   maxl[j]* B[j,l,u];
subject to Min_Prod {j in CROP, l in LOC}:
sum{p in WEEKP,u in TECH} Plant[p,j,l,u] >=  sum{u in TECH} minl[j]* B[j,l,u];
#####
#Resource Allocation
subject to Water_Allocation {(p,h) in WEEK6, z in ZON,j in CROP}:
sum{l in LOC, u in TECH: cidloc[l]==cidzone[z]} ( MinWReq[j,u] * Plant[p,j,l,u] ) <=
LRainRec[p,h,z]  +  WatAll[p,h,z,j];
# MicroHarvest restrictions
subject to Micro_harvest {h in WEEKH, j in CROP, z in ZON, u in TECH, l in
LOC:cidloc[l]==cidzone[z]}:    # Limit harvest by amount planted #
MicroHarv[h,j,l,u] =  sum{p in WEEKP:(p,h) in WEEK6} Plant[p,j,l,u] *YDist[p,h,j,z,u]
*  Yield[p,j,z,u];
subject to Shipment_L {h in WEEKH , j in CROP, q in QUAL,l in LOC, z in ZON:
cidloc[l]==cidzone[z]}:
SLZ[h,h,j,q,l,z] = sum{u in TECH:q==2 and cidloc[l]==cidzone[z]}
COL[h,j,q]*MicroHarv[h,j,l,u];
#Logistic restrictions
subject to Tot_packaging {h in WEEKH,j in CROP,q in QUAL,z in ZON}:
# Packaging quantity depends on amount harvested #
PACK[h,j,q,z] = sum{l in LOC,u in TECH:q==2 and cidloc[l]==cidzone[z]}
COL[h,j,q]*MicroHarv[h,j,l,u]/25;    #/ Weight[j]);
```

```
# Warehousing capacity at each ZON and DC
subject to Ware_Z_Cap {z in ZON, h in WEEKH}:
sum{j in CROP, q in QUAL: q==2} Invw[h,h,j,q,z]/25 <=WZ_Cap[z]+Add_War[z];
#Initial shipment to warehouse at each of the zones
subject to Initial_hold_W {h in WEEKH,j in CROP,q in QUAL,z in ZON}:
Invw[h,h,j,q,z] = sum{l in LOC: cidloc[l]==cidzone[z]} SLZ[h,h,j,q,l,z]- sum{d in DC}
SZD[h,h+TimeZD[z,d],j,q,z,d];
# Inventory at the warehouses
subject to Invent_W {(h,t) in WEEK1,j in CROP,q in QUAL,z in ZON: t>h}:
Invw[h,t,j,q,z] = Invw[h,t-1,j,q,z] - sum{d in DC} SZD[h,t+TimeZD[z,d],j,q,z,d];
#Initial shipment to warehouse at each of the zones
subject to Initial_hold_DC {h in WEEKH,j in CROP,q in QUAL,d in DC}:
Invd[h,h,j,q,d] = sum{z in ZON} SZD[h,h,j,q,z,d]- sum{i in CUST,m in MOD}
SDC[h,h,j,q,d,i,m];  # - sum{i in CUST,m in MOD} SDC[h,h+TimeDC[d,i,m],j,q,d,i,m];
# Inventory at the warehouses
subject to Invent_DC {(h,t) in WEEK1,j in CROP,q in QUAL,d in DC:t>h}:
Invd[h,t,j,q,d] = Invd[h,t-1,j,q,d] +sum{z in ZON} SZD[h,t,j,q,z,d] -sum{i in CUST,m in
MOD} SDC[h,t,j,q,d,i,m];
# Capacity at the warehouse
subject to Cap_warehouse {w in WARE, t in WEEKS}: sum{k in PROD,q in QUAL,h in
WEEKH: t>= h >= t-SL[k]} Invw[h,t,k,q,w]/Pallet[k]<= Wcap[w];
subject to Ship_const_Z {j in CROP, q in QUAL, z in ZON, d in DC}:
```

```
sum{(h,t) in WEEK1: t > h + TimeZD[z,d]} SZD[h,t,j,q,z,d] + sum{(h,t) in WEEK1: t < h
+ TimeZD[z,d]} SZD[h,t,j,q,z,d] =0;
subject to Ship_const_W {j in CROP, q in QUAL, z in ZON, i in CUST, m in MOD}:
sum{(h,t) in WEEK1: t > h + TimeZC[z,i,m]} SZC[h,t,j,q,z,i,m] + sum{(h,t) in WEEK1:t
< h + TimeZC[z,i,m]} SZC[h,t,j,q,z,i,m] =0;
subject to Ship_const_DC {j in CROP, q in QUAL, d in DC, i in CUST, m in MOD}:
sum{(h,t) in WEEK1: t > h + TimeDC[d,i,m]} SDC[h,t,j,q,d,i,m] + sum{(h,t) in WEEK1:
t < h + TimeDC[d,i,m]} SDC[h,t,j,q,d,i,m] =0;
subject to Ship_const_Z_Harv_Period:
sum{j in CROP, q in QUAL, d in DC,m in MOD, i in CUST, (h,t) in WEEK1: t >= 52}
SDC[h,t,j,q,d,i,m]  + sum{j in CROP, q in QUAL, d in DC,m in MOD, i in CUST, (h,t) in
WEEK1: t <= 13} SDC[h,t,j,q,d,i,m]<=0;
subject to Inv_Const_W:
sum{j in CROP, q in QUAL, z in ZON,(h,t) in WEEK1: t >= 52} Invw[h,t,j,q,z]  <=0;
subject to Inv_Const_DC:
sum{j in CROP, q in QUAL, d in DC, (h,t) in WEEK1: t >= 52} Invd[h,t,j,q,d]  <=0;
#Demand
subject to Max_Demand {j in CROP,i in CUST,t in WEEKS}:
sum{h in WEEKH, d in DC, q in QUAL,m in MOD: h + SL[j] >= t >= h and  (h,t) in
WEEK1 and q<=Qmin[i]} SDC[h,t,j,q,d,i,m]   <= MaxDem[t,j,i];
subject to Min_Demand {j in CROP,i in CUST,t in WEEKS}:
```

```
sum{h in WEEKH, d in DC, q in QUAL,m in MOD: h + SL[j] >= t >= h and  (h,t) in
WEEK1 and q<=Qmin[i]} SDC[h,t,j,q,d,i,m]   >= MinDem[t,j,i];
######
reset;
suffix dunbdd OUT;
option display1col 0;
option eexit -10000;
option csvdisplay_header 0;
option cplex_options 'mipdisplay=2';
option solver cplexamp;
model C:\Dropbox\PhD_Dissertation\AMPLcml\ID\Deterministic\V17\ID.mod;
#### Declaring Data from an Excel File
##Ranges (or sets)
table CUST IN "ODBC" "DSN=test" "SQL=SELECT * FROM
C:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\CUST_table.csv": CUST <-
[CUST], Qmin,LT;
table TECH IN "ODBC" "DSN=test" "SQL=SELECT * FROM
C:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\TECH.csv": TECH <- [TECH];
table CROP IN "ODBC" "DSN=test" "SQL=SELECT * FROM
C:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\CROP_table.csv": CROP <-
[CROP], minl,maxl,TraF,Cplant,Ccase,SL;
```

table LOC IN "ODBC" "DSN=test" "SQL=SELECT * FROM C:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\LOC_table.csv": LOC <-[LOC], cidloc, LAvail;

table ZON IN "ODBC" "DSN=test" "SQL=SELECT * FROM C:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\ZON_table.csv": ZON <-[ZON], cidzone, CTLZ, PZcap,WZ_Cap,Cwater,Cw;

table DC IN "ODBC" "DSN=test" "SQL=SELECT * FROM C:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\DC_table.csv": DC <-[DC],Cd;

table MODE IN "ODBC" "DSN=test" "SQL=SELECT * FROM C:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\MODE_table.csv": MOD <-[MODE];

table WEEKP IN "ODBC" "DSN=test" "SQL=SELECT * FROM C:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\WEEKP.csv": WEEKP <-[WEEKP];

table WEEKH IN "ODBC" "DSN=test" "SQL=SELECT * FROM C:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\WEEKH.csv": WEEKH <-[WEEKH];

table WEEKS IN "ODBC" "DSN=test" "SQL=SELECT * FROM C:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\WEEKS.csv": WEEKS <-[WEEKS];

table WEEK IN "ODBC" "DSN=test" "SQL=SELECT * FROM C:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\WEEK.csv": WEEK <- [WEEK];

```
table WEEK1 IN "ODBC" "DSN=test" "SQL=SELECT * FROM C:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\WEEK1.csv": WEEK1 <- [WEEKH,WEEKS];
table WEEK2 IN "ODBC" "DSN=test" "SQL=SELECT * FROM C:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\WEEK2.csv": WEEK2 <- [WEEKP, WEEKH,CROP,ZON,TECH];
table WEEK3 IN "ODBC" "DSN=test" "SQL=SELECT * FROM C:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\WEEK3.csv": WEEK3 <- [WEEKP, WEEKH,CROP,ZON];
table WEEK4 IN "ODBC" "DSN=test" "SQL=SELECT * FROM C:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\WEEK4.csv": WEEK4 <- [WEEKP, WEEKH,CROP];
table WEEK5 IN "ODBC" "DSN=test" "SQL=SELECT * FROM C:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\WEEK5.csv": WEEK5 <- [WEEKP, WEEKH,ZON];
table WEEK6 IN "ODBC" "DSN=test" "SQL=SELECT * FROM C:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\WEEK6.csv": WEEK6 <- [WEEKP, WEEKH];
table YIELD IN "ODBC" "DSN=test" "SQL=SELECT * FROM C:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\YIELDS_mean_table.csv": [WEEKP,CROP,ZON,TECH], Yield;
```

table YDIST IN "ODBC" "DSN=test" "SQL=SELECT * FROM C:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\YDist_table.csv": [WEEKP,WEEKH,CROP,ZON,TECH], YDist;

table DEM IN "ODBC" "DSN=test" "SQL=SELECT * FROM C:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\DEM_table.csv": [WEEKS,CROP,CUST], MaxDem, MinDem;

table CTZC IN "ODBC" "DSN=test" "SQL=SELECT * FROM C:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\CTZC_table.csv": [ZON,CUST,MOD], CTZC,TimeZC;

table CTZD IN "ODBC" "DSN=test" "SQL=SELECT * FROM C:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\CTZD_table.csv": [ZON,DC], CTZD,TimeZD;

table CTDC IN "ODBC" "DSN=test" "SQL=SELECT * FROM C:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\CTDC_table.csv": [DC,CUST,MOD], CTDC,TimeDC;

table CTECH IN "ODBC" "DSN=test" "SQL=SELECT * FROM C:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\CTECH_table.csv": [TECH,ZON], Ctech,Coper;

table COL IN "ODBC" "DSN=test" "SQL=SELECT * FROM C:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\COL_table.csv": [WEEKH,CROP,QUAL], COL;

```
table WATER IN "ODBC" "DSN=test" "SQL=SELECT * FROM C:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\WATER_REQ.csv":[CROP,TECH], MinWReq;
table ENVREC IN "ODBC" "DSN=test" "SQL=SELECT * FROM C:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\ENVREC_table.csv":[WEEKP,WEEKH,ZON],LRainRec;
table PRICE IN "ODBC" "DSN=test" "SQL=SELECT * FROM C:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\PRICE_mean_table.csv": [WEEKS,CROP,CUST], price;
table PARAMS IN "ODBC" "DSN=test" "SQL=SELECT * FROM C:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\PARAMS_table.csv": [], AvailCap;
##Data tables
# Loading Data to AMPL
read table CUST;
read table DC;
read table MODE;
read table CROP;
read table TECH;
read table WATER;
read table ENVREC;
read table WEEKP;
```

```
read table WEEKH;
read table WEEKS;
read table WEEK;
read table WEEK1;
read table WEEK2;
read table WEEK3;
read table WEEK4;
read table WEEK5;
read table WEEK6;
read table LOC;
read table ZON;
read table YDIST;
read table YIELD;
read table DEM;
read table PRICE;
read table CTZC;
read table CTZD;
read table CTDC;
read table CTECH;
read table COL;
read table PARAMS;
display minl;
```

```
# Solve Problem
problem Master:
#  Objective function
Yield_revenue,
# Constraints
TechInvL, Tot_land, Tot_land_Loc, TechType, M_Prod, Min_Prod, Water_Allocation,
Tot_packaging, Ship_const_Z_Harv_Period, Shipment_L, Ship_const_Z, Ship_const_W,
Ship_const_DC, Initial_hold_W, Invent_W, Initial_hold_DC, Invent_DC, Inv_Const_W ,
Inv_Const_DC, Max_Demand,Min_Demand, Micro_harvest, Ware_Z_Cap, Add_War,
## Decision variables
SLZ,SZD,SZC,SDC, PACK, Plant, WatAll,B, MicroHarv, Invw, Invd;
solve Master;
# Outputs the profits per location
let {j in CROP, d in DC, i in CUST}
Revenue_Profit [j,d,i] := sum {(h,t) in WEEK1, q in QUAL,m in MOD}
SDC[h,t,j,q,d,i,m]*price[t,j,i];
let {z in ZON,j in CROP}
Zone_Plant_Cost [z,j]:= sum{l in LOC,p in WEEKP, u in TECH: cidloc[l]==cidzone[z]}
Plant[p,j,l,u]*Cplant[j];
let {z in ZON,j in CROP}
Zone_Water_Cost [z,j]:= sum{(p,h) in WEEK6} WatAll[p,h,z,j]*Cwater[z];
let {z in ZON,j in CROP}
```

```
Zone_Invst_Cost [z,j]:= sum {p in WEEKP, l in LOC,u in TECH: cidloc[l]==cidzone[z]}
(Ctech[u,z]+Coper[u,z])*Plant[p,j,l,u];
let {j in CROP}
Logistic_Cost [j]:= sum {(h,t) in WEEK1, q in QUAL, z in ZON} Invw[h,t,j,q,z]*Cw[z]+
sum {(h,t) in WEEK1, q in QUAL, d in DC} Invd[h,t,j,q,d]*Cd[d]+
sum {(h,t) in WEEK1, q in QUAL,d in DC, i in CUST,m in MOD}
SDC[h,t,j,q,d,i,m]*CTDC[d,i,m] +        # From facility to customer
sum {(h,t) in WEEK1, q in QUAL,z in ZON,d in DC} SZD[h,t,j,q,z,d]*CTZD[z,d]+
# From facility to customer
sum {h in WEEKH, q in QUAL,l in LOC, z in ZON:cidloc[l]==cidzone[z]}
SLZ[h,h,j,q,l,z]*CTLZ[z];
let {z in ZON,j in CROP,u in TECH}
Tech_Selection [z,j,u]:= sum {l in LOC:cidloc[l]==cidzone[z]} B[j,l,u];
display _nvars;
display _ncons;
display _total_solve_time;
option send_statuses 0;
# Write output to the csv File
csvdisplay solve_result >> C:\Dropbox\PhD_Dissertation\AMPLcml\ID\AMPL_RUN\
deterministic_solve_result.csv;
csvdisplay Revenue_Profit >> C:\Dropbox\PhD_Dissertation\AMPLcml\ID\OUTPUT\
revenue_profit_deter.csv;
```

```
csvdisplay Zone_Plant_Cost >> C:\Dropbox\PhD_Dissertation\AMPLcml\ID\OUTPUT\
zon_plant_cost_deter.csv;
csvdisplay Zone_Water_Cost >> C:\Dropbox\PhD_Dissertation\AMPLcml\ID\OUTPUT\
zon_water_cost_deter.csv;
csvdisplay Zone_Invst_Cost >> C:\Dropbox\PhD_Dissertation\AMPLcml\ID\OUTPUT\
zon_invst_cost_deter.csv;
csvdisplay Logistic_Cost >> C:\Dropbox\PhD_Dissertation\AMPLcml\ID\OUTPUT\
logistic_cost_deter.csv;
csvdisplay Plant >> C:\Dropbox\PhD_Dissertation\AMPLcml\ID\OUTPUT\
plant_deter.csv;
csvdisplay MicroHarv >> C:\Dropbox\PhD_Dissertation\AMPLcml\ID\OUTPUT\
MicroHarv_deter.csv;
csvdisplay SZC >> C:\Dropbox\PhD_Dissertation\AMPLcml\ID\OUTPUT\szc_deter.csv;
csvdisplay SLZ >> C:\Dropbox\PhD_Dissertation\AMPLcml\ID\OUTPUT\slz_deter.csv;
csvdisplay Tech_Selection >> C:\Dropbox\PhD_Dissertation\AMPLcml\ID\OUTPUT\
b.csv;
csvdisplay SZD >> C:\Dropbox\PhD_Dissertation\AMPLcml\ID\OUTPUT\
szd_deter.csv;
csvdisplay SDC >> C:\Dropbox\PhD_Dissertation\AMPLcml\ID\OUTPUT\
sdc_deter.csv;
csvdisplay Invw >> C:\Dropbox\PhD_Dissertation\AMPLcml\ID\OUTPUT\
invw_deter.csv;
```

```
csvdisplay Invd >> C:\Dropbox\PhD_Dissertation\AMPLcml\ID\OUTPUT\invd_deter.csv;
table PLAN OUT "ODBC" "C:\Dropbox\PhD_Dissertation\AMPLcml\ID\OUTPUT\OUT_XLSX_FORMAT.xlsx":
{p in WEEKP, j in CROP, l in LOC, u in TECH} -> [WEEKP,CROP,LOC,TECH], Plant;
table MICRO OUT "ODBC" "C:\Dropbox\PhD_Dissertation\AMPLcml\ID\OUTPUT\OUT_XLSX_FORMAT.xlsx":
{(h,t) in WEEK1, j in CROP,q in QUAL, l in LOC,z in ZON:cidloc[l]== cidzone[z] } -> [WEEKH,WEEKS,CROP,QUAL,LOC,ZON], SLZ;
table HARV OUT "ODBC" "C:\Dropbox\PhD_Dissertation\AMPLcml\ID\OUTPUT\OUT_XLSX_FORMAT.xlsx":
{(p,h,j,z) in WEEK3} -> [WEEKP,WEEKH,CROP,ZON], ZoneHarv;
table PAC OUT "ODBC" "C:\Dropbox\PhD_Dissertation\AMPLcml\ID\OUTPUT\OUT_XLSX_FORMAT.xlsx":
{h in WEEKH,j in CROP,q in QUAL,z in ZON: q==2} -> [WEEKH,CROP,QUAL,ZON], PACK;
table SHZC OUT "ODBC" "C:\Dropbox\PhD_Dissertation\AMPLcml\ID\OUTPUT\OUT_XLSX_FORMAT.xlsx":
{(h,t) in WEEK1,j in CROP,q in QUAL, z in ZON, i in CUST,m in MOD: SZC[h,t,j,q,z,i,m]>0} -> [WEEKH, WEEKS,CROP,QUAL,ZON,CUST,MOD], SZC;
table SHDC OUT "ODBC" "C:\Dropbox\PhD_Dissertation\AMPLcml\ID\OUTPUT\OUT_XLSX_FORMAT.xlsx":
```

```
{(h,t) in WEEK1,j in CROP,q in QUAL, d in DC, i in CUST,m in MOD:
SDC[h,t,j,q,d,i,m]>0}  -> [WEEKH, WEEKS,CROP,QUAL,DC,CUST,MOD], SDC;
table SHZD OUT "ODBC"
"C:\Dropbox\PhD_Dissertation\AMPLcml\ID\OUTPUT\OUT_XLSX_FORMAT.xlsx":
{(h,t) in WEEK1,j in CROP,q in QUAL, z in ZON, d in DC: SZD[h,t,j,q,z,d]>0}  ->
[WEEKH, WEEKS,CROP,QUAL,ZON,DC], SZD;
table INVW OUT "ODBC"
"C:\Dropbox\PhD_Dissertation\AMPLcml\ID\OUTPUT\OUT_XLSX_FORMAT.xlsx":
{(h,t) in WEEK1,j in CROP,q in QUAL, z in ZON:Invw[h,t,j,q,z]>0}  ->
[WEEKH,WEEKS,CROP,QUAL,ZON], Invw;
table INVD OUT "ODBC"
"C:\Dropbox\PhD_Dissertation\AMPLcml\ID\OUTPUT\OUT_XLSX_FORMAT.xlsx":
{(h,t) in WEEK1,j in CROP,q in QUAL, d in DC: Invd[h,t,j,q,d]>0} ->[WEEKH,
WEEKS,CROP,QUAL,DC], Invd;
```

APPENDIX C

MULTI-CUT L-SHAPED FORMULATION

```
### Model

############################################################

# Data sets #

set WEEKP ordered;  # weeks of the planning period

set WEEKH ordered;  # weeks of the harvesting period

set CROP;  # Crops for planting

set CUST; # Customers

set TECH;                                   # Growing technologies

set ZON;                       # Zones to be considered

set LOC;# Locations available for planting cropss

set SCEN;

set DC;

set MOD;

#####

set WEEK ordered;

set WEEKS ordered;

#Additional necessary sets to facilitate indexing

set WEEK1 within {WEEKH, WEEKS};

set WEEK2 within {WEEKP, WEEKH,CROP,ZON,TECH};

set WEEK3 within {WEEKP, WEEKH,CROP,ZON};

set WEEK4 within {WEEKP, WEEKH,CROP};

set WEEK5 within {WEEKP, WEEKH,ZON};
```

```
set WEEK6 within {WEEKP, WEEKH};

set QUAL:= 2..2;                        # Color characteristic of products

set ITER:= 1..100;

param NbScen >= 0;

########

# Parameter definition #

# Production/Yield-Related

param YDist {WEEKP,WEEKH,CROP,ZON,TECH} >= 0;     # Expected YDist per crop

param yield{WEEKP,CROP,ZON,TECH} >= 0;

param LRainRec{WEEKP,WEEKH,ZON} >= 0;

param lrainrec{WEEKP,WEEKH,ZON} default 0;

param Salv {WEEKP,WEEKH,CROP,ZON,TECH} >=0;  # Expected waste of crops

param MaxDem {WEEKS,CROP,CUST} >= 0;            # Maximum demand

param MinDem {WEEKS,CROP,CUST} >= 0;          # Demand from customer

param Qmin {CUST}>=0;                           # Quality demanded by customer i

param COL {WEEKH,CROP,QUAL}>=0;                   # Binary indicator parameter

param Maxi{LOC} >=0;# Maximum amount of hectares to plant in a given period

param minl{CROP} >=0;                           # Minimum amount to plant per crop

param maxl{CROP} >=0;                           # Maximum amount to plant per crop

param Ctech{TECH, ZON} >=0;                        # Cost of technology u for location l

param Coper{TECH, ZON} >= 0;  # Cost per acre for operating technology u in location l

param Cplant{CROP} >=0;
```

```
param Cwater{ZON}  >=0;                          # Cost of water per location l

param cidloc {LOC} symbolic;

param cidzone{ZON} symbolic;

# Environment-Related

param MinWReq {CROP,TECH} >=0;        # Water required per acre of crop

param LAvail {LOC} >=0;                  # Available hectares for planting at location l

# Logistics Related

param SL {CROP} >=0;                         # Shelf life of product k

param LT {CUST} >=0;   # Lead time required by the customer

param Weight {CROP} >=0; # Quantity in required of crop j to form a case of product k

param PZcap {ZON} >=0;            # Capacity of the packaging facility for a time period

param TimeZC {ZON,CUST,MOD} >=0;     # Time of transportation from zone to market

param TimeDC {DC,CUST,MOD} >=0;      # Time of transportation from DC to customer

param TimeZD {ZON,DC} >=0;                 # Time of transportation from zone z to DC d

param CTZD{ZON,DC} >=0;# Cost of transportation from zone z to customer c

param CTDC{DC,CUST,MOD} >=0;#Cost of transportation from zone z to customer c

param CTLZ{ZON} >=0;

param Cw {ZON} >= 0; # Cost of warehouse

param Cd {DC} >= 0;

param Ccase{CROP} >=0;                           # Cost for packing case for crop j

param TraF{CROP}  >=0;

param WZ_Cap{ZON} >= 0;
```

```
# Market Related

param price {WEEKS,CROP,CUST} >= 0;   # price for customer i per week t from crop j

# Additional parameters

param cur_scen >=0;                          # Investment available for the season

param cur_CUT >=0;

# Water available

param AvailCap >=0;                          # Capital available for technology investments

# Helper parameters for outputs

param PZC {CROP,LOC,ZON,TECH} ;

param nCUT >= 0 integer;# Counter of optimality cuts

param fCUT >= 0 integer;

param Type_cut >= 0 integer;

param count{ITER} default 0;

param theta_k{SCEN} default 0;

param break_check default 0;

param counter_last default 0;

set CUTS:=1..nCUT;

param value default 0;

####### Stochastic parameters and shadow prices

param obj_cotheta_k{SCEN} >=0;

param Prob{SCEN} >=0;

param yield_sub {SCEN,CUTS,WEEKP,CROP,ZON,TECH} >= 0;
```

param plant_master {WEEKP,CROP,LOC,TECH} >= 0;

param Zone_harvest_dual {SCEN,1..nCUT,WEEK3};

param Zone_harvest_ray {SCEN,1..fCUT,WEEK3};

param Tot_packaging_dual {SCEN,1..nCUT, WEEKH,CROP, q in QUAL,ZON};

param Tot_packaging_ray {SCEN,1..fCUT, WEEKH,CROP, q in QUAL, ZON};

param Ship_const_dual {SCEN,1..nCUT,CROP, q in QUAL,ZON,CUST};

param Ship_const_ray {SCEN,1..fCUT,CROP, q in QUAL, ZON,CUST};

param Max_Demand_dual {SCEN,1..nCUT,CROP,CUST,WEEKS};

param Max_Demand_ray {SCEN,1..fCUT,CROP,CUST,WEEKS};

param Min_Demand_dual {SCEN,1..nCUT,CROP,CUST,WEEKS};

param Min_Demand_ray {SCEN,1..fCUT,CROP,CUST,WEEKS};

param Ware_Z_Cap_dual {SCEN,1..nCUT,ZON,WEEKH};

param Ware_Z_Cap_ray {SCEN,1..fCUT,ZON,WEEKH};

param Shipment_L_dual {SCEN, 1..nCUT, WEEKH, CROP,q in QUAL,l in LOC,z in ZON: cidloc[l]==cidzone[z]};

param Shipment_L_FIELD_ray {SCEN, 1..fCUT, WEEKH, CROP,q in QUAL, l in LOC,z in ZON: cidloc[l]==cidzone[z]};

param Ship_const_Z_dual {SCEN,1..nCUT,CROP,QUAL, ZON,DC};

param Ship_const_W_dual {SCEN,1..nCUT,CROP,QUAL, ZON,CUST,MOD};

param Ship_const_DC_dual {SCEN,1..nCUT,CROP,QUAL, DC,CUST,MOD};

param Invent_W_dual {SCEN,1..nCUT,(h,t) in WEEK1,CROP,q in QUAL,ZON:t>h};

param Invent_W_ray {SCEN,1..fCUT,(h,t) in WEEK1,CROP,q in QUAL,ZON:t>h};

```
param Invent_DC_dual {SCEN,1..nCUT,(h,t) in WEEK1,CROP,q in QUAL, DC:t>h};
param Invent_DC_ray {SCEN,1..fCUT,(h,t) in WEEK1,CROP,q in QUAL, DC:t>h};


param Initial_hold_W_dual {SCEN,1..nCUT,WEEKH,CROP,q in QUAL,ZON};
param Initial_hold_W_ray {SCEN,1..fCUT, WEEKH,CROP,q in QUAL,ZON};
param Initial_hold_DC_dual {SCEN,1..nCUT,WEEKH,CROP,q in QUAL,DC};
param Initial_hold_DC_ray {SCEN,1..fCUT,WEEKH,CROP,q in QUAL,DC};
param Micro_harvest_dual{SCEN,1..nCUT,h in WEEKH, j in CROP, z in ZON, u in
TECH, l in LOC:cidloc[l]==cidzone[z]};
param Micro_harvest_ray{SCEN,1..fCUT,h in WEEKH, j in CROP, z in ZON, u in TECH,
l in LOC:cidloc[l]==cidzone[z]};
param Water_Allocation_dual {SCEN,1..nCUT, WEEK5};
##########
param VMicroHarv_Output {SCEN,1..nCUT,h in WEEKH, j in CROP, l in LOC,u in
TECH};
param VPlant_Output {1..nCUT,WEEKP,CROP,LOC,TECH};
param VTheta_Output {1..nCUT,SCEN};
param VB_Output {1..nCUT,CROP,LOC,TECH};
########## Variable Definition
#Planting and resource decisions
var VPlant {WEEKP,CROP,LOC,TECH} >=0; # planting crop j, in period p at location l
var VMicroHarv {WEEKH,CROP,LOC,TECH} >=0;     # Harvest (pounds)
```

```
var VZoneHarv {(p,h,j,z) in WEEK3}  >=0;
var SumPerishable>=0;
var VWatAll {WEEK5} >=0;  # Wtr qty allocated to location l for crop j using tech u
var VB {CROP,LOC,TECH} >= 0 binary;  # Binary decision to plant crop j in l using
#technology u
#Logistic related variables
var VPACK {WEEKH,CROP,QUAL,ZON} >=0;              # Quantity of crop j packed
var VSLZ {WEEK1,CROP,QUAL,l in LOC,z in ZON:cidloc[l]==cidzone[z]} >=0;
var VSZD {WEEK1,CROP,QUAL,ZON,DC} >=0;
var VSDC {WEEK1,CROP,QUAL,DC,CUST,MOD} >= 0;
var VInvw {WEEK1,CROP,QUAL,ZON} >=0;  # Invtory in warehouse within each zone
var VInvd {WEEK1,CROP,QUAL,DC} >=0;  # Inventory at the DC
### Softening of constraints
var  Add_War {ZON} >= 0;
var theta{SCEN};
#####################
# MASTER PROBLEM #
# Objective function
maximize first_stage:
-sum{p in WEEKP,j in CROP,l in LOC,u in TECH} (VPlant[p,j,l,u]*Cplant[j])
# Planting Costs (at this point assuming that it does not vary based on technology
# Technology costs
```

```
-sum {p in WEEKP, z in ZON, u in TECH, j in CROP, l in LOC: cidloc[l]==cidzone[z]}
(Coper[u,z]+Ctech[u,z])*VPlant[p,j,l,u]   # Becomes the technology costs
-sum {s in SCEN} theta[s];
# Constraints
subject to TechInvL:      # Investment in Technologies is limited by available capital
sum{z in ZON, j in CROP, l in LOC,  u in TECH: cidloc[l]== cidzone[z]}
VB[j,l,u]*Ctech[u,z] <= AvailCap;
subject to Tot_land {l in LOC, j in CROP, u in TECH}:
sum{p in WEEKP} VPlant[p,j,l,u] <= LAvail[l]*VB[j,l,u];
subject to Tot_land_Loc {l in LOC}:
sum{p in WEEKP,j in CROP, u in TECH} VPlant[p,j,l,u] <= LAvail[l];
subject to TechType{j in CROP}:
sum{u in TECH,l in LOC} VB[j,l,u] <=5;
subject to M_Prod {p in WEEKP, j in CROP, l in LOC, u in TECH}:
VPlant[p,j,l,u] <= maxl[j]*  VB[j,l,u];
subject to Min_Prod {j in CROP, l in LOC}:
sum{p in WEEKP,u in TECH} VPlant[p,j,l,u] >= sum{u in TECH} minl[j]*VB[j,l,u];
##### Optimality cuts
subject to Cut_Defn {s in SCEN, c in 1..nCUT}:
theta[s] >=
```

- sum{(p,h) in WEEK6, j in CROP, z in ZON, u in TECH, l in LOC:cidloc[l]==cidzone[z]} Prob[s]*Micro_harvest_dual[s,c,h,j,z,u,l]*(VPlant[p,j,l,u]*YDist[p,h,j,z,u]*yield_sub[s,c, p,j,z,u])

- sum {(p,h,z) in WEEK5} Prob[s]*Water_Allocation_dual[s,c,p,h,z]*sum{l in LOC, u in TECH, j in CROP: cidloc[l]==cidzone[z]} (MinWReq[j,u] * VPlant[p,j,l,u])

+ sum{(p,h,z) in WEEK5} Prob[s]*Water_Allocation_dual[s,c,p,h,z]*-1*LRainRec[p,h,z]

+ sum{h in WEEKH, j in CROP, q in QUAL,l in LOC, z in ZON: cidloc[l]==cidzone[z]}Prob[s]*Shipment_L_dual[s,c,h,j,q,l,z]*COL[h,j,q]

+ sum{h in WEEKH,j in CROP,q in QUAL,z in ZON}Prob[s]*Initial_hold_W_dual[s,c,h,j,q,z]

+ sum{(h,t) in WEEK1,j in CROP,q in QUAL,z in ZON: t>h}Prob[s]*Invent_W_dual[s,c,h,t,j,q,z]

+ sum{h in WEEKH,j in CROP,q in QUAL,d in DC}Prob[s]*Initial_hold_DC_dual[s,c,h,j,q,d]

+ sum{(h,t) in WEEK1,j in CROP,q in QUAL,d in DC: t>h}Prob[s]*Invent_DC_dual[s,c,h,t,j,q,d]

+ sum{h in WEEKH,j in CROP,q in QUAL,z in ZON}Prob[s]*Tot_packaging_dual[s,c,h,j,q,z]*COL[h,j,q]

+ sum{j in CROP,i in CUST,t in WEEKS} Prob[s]*Max_Demand_dual[s,c,j,i,t]*MaxDem[t,j,i]

+ sum{j in CROP,i in CUST,t in WEEKS} Prob[s]*Min_Demand_dual[s,c,j,i,t]*-1*MinDem[t,j,i]

```
+ sum{z in ZON, h in WEEKH} Prob[s]*Ware_Z_Cap_dual[s,c,z,h]*WZ_Cap[z]

+ sum{j in CROP,q in QUAL,z in ZON, d in DC} Prob[s]*Ship_const_Z_dual[s,c,j,q,z,d]

+ sum {j in CROP,q in QUAL,z in ZON, i in CUST, m in MOD}
Prob[s]*Ship_const_W_dual[s,c,j,q,z,i,m]

+ sum {j in CROP,q in QUAL,d in DC, i in CUST, m in MOD}
Prob[s]*Ship_const_DC_dual[s,c,j,q,d,i,m]

#################

# SUBPROBLEM

# Objective function

maximize second_stage:

(

# Market revenues

sum {(h,t) in WEEK1, j in CROP, q in QUAL,i in CUST,d in DC, m in MOD}
VSDC[h,t,j,q,d,i,m]*price[t,j,i]*0.5   ##Selling to customer directly from packing facility

-sum{z in ZON,h in WEEKH,j in CROP, q in QUAL} (VPACK[h,j,q,z]/25*Ccase[j])

# Packing costs

-sum{(h,t) in WEEK1, j in CROP, q in QUAL, z in ZON} VInvw[h,t,j,q,z]*Cw[z]

-sum{(h,t) in WEEK1, j in CROP, q in QUAL, d in DC} VInvd[h,t,j,q,d]*Cd[d]

-sum{ (p,h,z) in WEEK5} (VWatAll[p,h,z]*Cwater[z])

-sum {(h,t) in WEEK1, j in CROP,q in QUAL,d in DC, i in CUST,m in MOD}
VSDC[h,t,j,q,d,i,m]*CTDC[d,i,m]          # From facility to customer
```

```
-sum {(h,t) in WEEK1, j in CROP,q in QUAL,z in ZON,d in DC}
VSZD[h,t,j,q,z,d]*CTZD[z,d]          # From facility to customer
-sum {(h,t) in WEEK1, j in CROP,q in QUAL,l in LOC, z in ZON:cidloc[l]==cidzone[z]}
VSLZ[h,t,j,q,l,z]*CTLZ[z]
-sum{z in ZON} 10000000*Add_War[z]
) ;
#######
# Constraints
##Resource Allocation
subject to Water_Allocation {(p,h,z) in WEEK5}:
VWatAll[p,h,z] >= -LRainRec[p,h,z] + sum{l in LOC, u in TECH, j in CROP:
cidloc[l]==cidzone[z]} ( MinWReq[j,u] * plant_master[p,j,l,u] );
subject to Micro_harvest {h in WEEKH, j in CROP, z in ZON, u in TECH, l in
LOC:cidloc[l]==cidzone[z]}:   # Limit harvest by amount planted #
VMicroHarv[h,j,l,u] =  sum{p in WEEKP:(p,h) in WEEK6} plant_master[p,j,l,u]
*YDist[p,h,j,z,u] * yield[p,j,z,u];
#Logistic restrictions
subject to Shipment_L {h in WEEKH, j in CROP, q in QUAL,l in LOC, z in ZON:
cidloc[l]==cidzone[z]}:
VSLZ[h,h,j,q,l,z] = sum{u in TECH:q==2 and cidloc[l]==cidzone[z]}
COL[h,j,q]*VMicroHarv[h,j,l,u];
#Logistic restrictions
```

subject to Tot_packaging {h in WEEKH,j in CROP,q in QUAL,z in ZON}:

VPACK[h,j,q,z] = sum{l in LOC,u in TECH:q==2 and cidloc[l]==cidzone[z]} COL[h,j,q]*VMicroHarv[h,j,l,u]/25;

subject to Ware_Z_Cap {z in ZON, h in WEEKH}:

sum{j in CROP, q in QUAL: q==2} VInvw[h,h,j,q,z]/25 <=WZ_Cap[z]+Add_War[z];

#Initial shipment to warehouse at each of the zones

subject to Initial_hold_W {h in WEEKH,j in CROP,q in QUAL,z in ZON}:

VInvw[h,h,j,q,z] = sum{l in LOC: cidloc[l]==cidzone[z]} VSLZ[h,h,j,q,l,z]- sum{d in DC} VSZD[h,h+TimeZD[z,d],j,q,z,d]; # -sum{d in DC} SZD[h,h+TimeZD[z,d],j,q,z,d]; # - sum{i in CUST} SZC[h,h+TimeZC[z,i],j,q,z,i] - sum{d in DC} SZD[h,h+TimeZD[z,d],j,q,z,d];

Inventory at the warehouses

subject to Invent_W {(h,t) in WEEK1,j in CROP,q in QUAL,z in ZON: t>h}:

VInvw[h,t,j,q,z] = VInvw[h,t-1,j,q,z] - sum{d in DC} VSZD[h,t+TimeZD[z,d],j,q,z,d]; # - sum{i in CUST,m in MOD} SZC[h,h+TimeZC[z,i,m],j,q,z,i,m];

#Initial shipment to warehouse at each of the zones

subject to Initial_hold_DC {h in WEEKH,j in CROP,q in QUAL,d in DC}:

VInvd[h,h,j,q,d] = sum{z in ZON} VSZD[h,h,j,q,z,d]- sum{i in CUST,m in MOD} VSDC[h,h,j,q,d,i,m]; # - sum{i in CUST,m in MOD} SDC[h,h+TimeDC[d,i,m],j,q,d,i,m];

Inventory at the warehouses

subject to Invent_DC {(h,t) in WEEK1,j in CROP,q in QUAL,d in DC:t>h}:

```
VInvd[h,t,j,q,d] = VInvd[h,t-1,j,q,d] +sum{z in ZON} VSZD[h,t,j,q,z,d] -sum{i in CUST,m in MOD} VSDC[h,t,j,q,d,i,m];
# Capacity at the warehouse
subject to Ship_const_Z {j in CROP, q in QUAL, z in ZON, d in DC}:
sum{h in WEEKH,t in WEEKS: t > h + TimeZD[z,d] and (h,t) in WEEK1} VSZD[h,t,j,q,z,d] + sum{h in WEEKH,t in WEEKS: t < h + TimeZD[z,d] and (h,t) in WEEK1} VSZD[h,t,j,q,z,d] <=0;
subject to Ship_const_W {j in CROP, q in QUAL, z in ZON, i in CUST, m in MOD}:
sum{h in WEEKH,t in WEEKS: t > h + TimeZC[z,i,m] and (h,t) in WEEK1} VSZC[h,t,j,q,z,i,m] + sum{h in WEEKH,t in WEEKS: t < h + TimeZC[z,i,m] and (h,t) in WEEK1} VSZC[h,t,j,q,z,i,m] <=0;
subject to Ship_const_DC {j in CROP, q in QUAL, d in DC, i in CUST, m in MOD}:
sum{h in WEEKH,t in WEEKS: t > h + TimeDC[d,i,m] and (h,t) in WEEK1} VSDC[h,t,j,q,d,i,m] + sum{h in WEEKH,t in WEEKS: t < h + TimeDC[d,i,m] and (h,t) in WEEK1} VSDC[h,t,j,q,d,i,m] <=0;
subject to Inv_Const_W:
sum{j in CROP, q in QUAL, z in ZON,(h,t) in WEEK1: t > 53} VInvw[h,t,j,q,z]  <=0;
subject to Inv_Const_DC:
sum{j in CROP, q in QUAL, d in DC, (h,t) in WEEK1: t > 53} VInvd[h,t,j,q,d]  <=0;
#Demand
subject to Max_Demand {j in CROP,i in CUST,t in WEEKS}:
```

```
sum{h in WEEKH, d in DC, q in QUAL,m in MOD: h + SL[j] >= t >= h and q<=Qmin[i]
and (h,t) in WEEK1} VSDC[h,t,j,q,d,i,m]   <= MaxDem[t,j,i];
subject to Min_Demand {j in CROP,i in CUST,t in WEEKS}:
sum{h in WEEKH, d in DC, q in QUAL,m in MOD: h + SL[j] >= t >= h and q<=Qmin[i]
and (h,t) in WEEK1} VSDC[h,t,j,q,d,i,m]   >= MinDem[t,j,i];
#####
### SUBPROBLEM
######
#################
# SUBPROBLEM
# Objective function

maximize second_stage:
(
# Market revenues
sum {(h,t) in WEEK1, j in CROP, q in QUAL,i in CUST,d in DC, m in MOD}
VSDC[h,t,j,q,d,i,m]*price[t,j,i]   ##Selling to customer directly from packing facility
-sum{z in ZON,h in WEEKH,j in CROP, q in QUAL} (VPACK[h,j,q,z]/25*Ccase[j])
# Packing costs
-sum{(h,t) in WEEK1, j in CROP, q in QUAL, z in ZON} VInvw[h,t,j,q,z]*Cw[z]
-sum{(h,t) in WEEK1, j in CROP, q in QUAL, d in DC} VInvd[h,t,j,q,d]*Cd[d]
-sum{ (p,h,z) in WEEK5} (VWatAll[p,h,z]*Cwater[z])
```

Transportation costs

-sum {(h,t) in WEEK1, j in CROP,q in QUAL,d in DC, i in CUST,m in MOD} VSDC[h,t,j,q,d,i,m]*CTDC[d,i,m]   # From facility to customer

-sum {(h,t) in WEEK1, j in CROP,q in QUAL,z in ZON,d in DC} VSZD[h,t,j,q,z,d]*CTZD[z,d]   # From facility to customer

-sum {(h,t) in WEEK1, j in CROP,q in QUAL,l in LOC, z in ZON:cidloc[l]==cidzone[z]} VSLZ[h,t,j,q,l,z]*CTLZ[z]

Keep feasibility

-sum{z in ZON} 10000000*Add_War[z]

);

#######

Constraints

##Resource Allocation

subject to Water_Allocation {(p,h,z) in WEEK5}:   # Water amount needed for the Yield period in location for each crop and technology (FOR NOW TREAT AS DETERMINISTIC)

VWatAll[p,h,z] >= -LRainRec[p,h,z] + sum{l in LOC, u in TECH, j in CROP: cidloc[l]==cidzone[z]} (MinWReq[j,u] * plant_master[p,j,l,u]);

Obtains the harvested product at a micro level

subject to Micro_harvest {h in WEEKH, j in CROP, z in ZON, u in TECH, l in LOC:cidloc[l]==cidzone[z]}:   # Limit harvest by amount planted #

VMicroHarv[h,j,l,u] = sum{p in WEEKP:(p,h) in WEEK6} plant_master[p,j,l,u] *YDist[p,h,j,z,u] * yield[p,j,z,u];

#Logistic restrictions

subject to Shipment_L {h in WEEKH, j in CROP, q in QUAL,l in LOC, z in ZON: cidloc[l]==cidzone[z]}:

VSLZ[h,h,j,q,l,z] = sum{u in TECH:q==2 and cidloc[l]==cidzone[z]} COL[h,j,q]*VMicroHarv[h,j,l,u];

#Logistic restrictions

subject to Tot_packaging {h in WEEKH,j in CROP,q in QUAL,z in ZON}: # Packaging quantity depends on amount harvested #

VPACK[h,j,q,z] = sum{l in LOC,u in TECH:q==2 and cidloc[l]==cidzone[z]} COL[h,j,q]*VMicroHarv[h,j,l,u]/25;

subject to Ware_Z_Cap {z in ZON, h in WEEKH}:

sum{j in CROP, q in QUAL: q==2} VInvw[h,h,j,q,z]/25 <=WZ_Cap[z]+Add_War[z];

#Initial shipment to warehouse at each of the zones

subject to Initial_hold_W {h in WEEKH,j in CROP,q in QUAL,z in ZON}:

VInvw[h,h,j,q,z] = sum{l in LOC: cidloc[l]==cidzone[z]} VSLZ[h,h,j,q,l,z]- sum{d in DC} VSZD[h,h+TimeZD[z,d],j,q,z,d]; # -sum{d in DC} SZD[h,h+TimeZD[z,d],j,q,z,d];

Inventory at the warehouses

subject to Invent_W {(h,t) in WEEK1,j in CROP,q in QUAL,z in ZON: t>h}:

VInvw[h,t,j,q,z] = VInvw[h,t-1,j,q,z] - sum{d in DC} VSZD[h,t+TimeZD[z,d],j,q,z,d]; # - sum{i in CUST,m in MOD} SZC[h,h+TimeZC[z,i,m],j,q,z,i,m];

#Initial shipment to warehouse at each of the zones

subject to Initial_hold_DC {h in WEEKH,j in CROP,q in QUAL,d in DC}:

VInvd[h,h,j,q,d] = sum{z in ZON} VSZD[h,h,j,q,z,d]- sum{i in CUST,m in MOD} VSDC[h,h,j,q,d,i,m];

Inventory at the warehouses

subject to Invent_DC {(h,t) in WEEK1,j in CROP,q in QUAL,d in DC:t>h}:

VInvd[h,t,j,q,d] = VInvd[h,t-1,j,q,d] +sum{z in ZON} VSZD[h,t,j,q,z,d] -sum{i in CUST,m in MOD} VSDC[h,t,j,q,d,i,m];

Capacity at the warehouse

subject to Ship_const_Z {j in CROP, q in QUAL, z in ZON, d in DC}:

sum{h in WEEKH,t in WEEKS: t > h + TimeZD[z,d] and (h,t) in WEEK1} VSZD[h,t,j,q,z,d] + sum{h in WEEKH,t in WEEKS: t < h + TimeZD[z,d] and (h,t) in WEEK1} VSZD[h,t,j,q,z,d] <=0;

subject to Ship_const_W {j in CROP, q in QUAL, z in ZON, i in CUST, m in MOD}:

sum{h in WEEKH,t in WEEKS: t > h + TimeZC[z,i,m] and (h,t) in WEEK1} VSZC[h,t,j,q,z,i,m] + sum{h in WEEKH,t in WEEKS: t < h + TimeZC[z,i,m] and (h,t) in WEEK1} VSZC[h,t,j,q,z,i,m] <=0;

subject to Ship_const_DC {j in CROP, q in QUAL, d in DC, i in CUST, m in MOD}:

sum{h in WEEKH,t in WEEKS: t > h + TimeDC[d,i,m] and (h,t) in WEEK1} VSDC[h,t,j,q,d,i,m] + sum{h in WEEKH,t in WEEKS: t < h + TimeDC[d,i,m] and (h,t) in WEEK1} VSDC[h,t,j,q,d,i,m] <=0;

subject to Inv_Const_W:

sum{j in CROP, q in QUAL, z in ZON,(h,t) in WEEK1: t > 53} VInvw[h,t,j,q,z] <=0;

subject to Inv_Const_DC:

sum{j in CROP, q in QUAL, d in DC, (h,t) in WEEK1: t > 53} VInvd[h,t,j,q,d] <=0;

#Demand

subject to Max_Demand {j in CROP,i in CUST,t in WEEKS}: # Customer demand is met either through shipping from field to customer or from warehouse to customer # Modified. Took out Warehouse, only considered DC

#option solver cplex; ##

reset;

suffix dunbdd OUT;

option display1col 0;

option eexit -10000;

option solver cplexamp;

model D:\Dropbox\PhD_Dissertation\AMPLcml\ID\Stochastic\BD_V19\ID.mod;

Declaring Data from an Excel File

##Ranges (or sets)

table CUST IN "ODBC" "DSN=test" "SQL=SELECT * FROM D:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\CUST_table.csv": CUST <- [CUST], Qmin,LT;

table TECH IN "ODBC" "DSN=test" "SQL=SELECT * FROM D:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\TECH.csv": TECH <- [TECH];

```
table CROP IN "ODBC" "DSN=test" "SQL=SELECT * FROM D:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\CROP_table.csv": CROP <-[CROP], minl,maxl,TraF,Cplant,Ccase,SL;
table LOC IN "ODBC" "DSN=test" "SQL=SELECT * FROM D:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\LOC_table.csv": LOC <-[LOC], cidloc, LAvail;
table ZON IN "ODBC" "DSN=test" "SQL=SELECT * FROM D:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\ZON_table.csv": ZON <-[ZON], cidzone, CTLZ, PZcap,WZ_Cap,Cwater,Cw;
table DC IN "ODBC" "DSN=test" "SQL=SELECT * FROM D:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\DC_table.csv": DC <-[DC],Cd;
table MODE IN "ODBC" "DSN=test" "SQL=SELECT * FROM D:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\MODE_table.csv": MOD <-[MODE];
table WEEKP IN "ODBC" "DSN=test" "SQL=SELECT * FROM D:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\WEEKP.csv": WEEKP <-[WEEKP];
table WEEKH IN "ODBC" "DSN=test" "SQL=SELECT * FROM D:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\WEEKH.csv": WEEKH <-[WEEKH];
```

```
table WEEKS IN "ODBC" "DSN=test" "SQL=SELECT * FROM D:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\WEEKS.csv": WEEKS <- [WEEKS];
table WEEK IN "ODBC" "DSN=test" "SQL=SELECT * FROM D:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\WEEK.csv": WEEK <- [WEEK];
table WEEK1 IN "ODBC" "DSN=test" "SQL=SELECT * FROM D:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\WEEK1.csv": WEEK1 <- [WEEKH,WEEKS];
table WEEK2 IN "ODBC" "DSN=test" "SQL=SELECT * FROM D:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\WEEK2.csv": WEEK2 <- [WEEKP, WEEKH,CROP,ZON,TECH];
table WEEK3 IN "ODBC" "DSN=test" "SQL=SELECT * FROM D:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\WEEK3.csv": WEEK3 <- [WEEKP, WEEKH,CROP,ZON];
table WEEK4 IN "ODBC" "DSN=test" "SQL=SELECT * FROM D:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\WEEK4.csv": WEEK4 <- [WEEKP, WEEKH,CROP];
table WEEK5 IN "ODBC" "DSN=test" "SQL=SELECT * FROM D:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\WEEK5.csv": WEEK5 <- [WEEKP, WEEKH,ZON];
```

table WEEK6 IN "ODBC" "DSN=test" "SQL=SELECT * FROM D:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\WEEK6.csv": WEEK6 <- [WEEKP, WEEKH];

table YDIST IN "ODBC" "DSN=test" "SQL=SELECT * FROM D:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\YDist_table.csv": [WEEKP,WEEKH,CROP,ZON,TECH], YDist;

table DEM IN "ODBC" "DSN=test" "SQL=SELECT * FROM D:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\DEM_table.csv": [WEEKS,CROP,CUST], MaxDem, MinDem;

table CTZC IN "ODBC" "DSN=test" "SQL=SELECT * FROM D:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\CTZC_table.csv": [ZON,CUST,MOD], CTZC,TimeZC;

table CTZD IN "ODBC" "DSN=test" "SQL=SELECT * FROM D:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\CTZD_table.csv": [ZON,DC], CTZD,TimeZD;

table CTDC IN "ODBC" "DSN=test" "SQL=SELECT * FROM D:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\CTDC_table.csv": [DC,CUST,MOD], CTDC,TimeDC;

table CTECH IN "ODBC" "DSN=test" "SQL=SELECT * FROM D:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\CTECH_table.csv": [TECH,ZON], Ctech,Coper;

```
table COL IN "ODBC" "DSN=test" "SQL=SELECT * FROM D:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\COL_table.csv":
[WEEKH,CROP,QUAL], COL;
table WATER IN "ODBC" "DSN=test" "SQL=SELECT * FROM D:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\WATER_REQ.csv":[CROP,TECH], MinWReq;
table PARAMS IN "ODBC" "DSN=test" "SQL=SELECT * FROM D:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\PARAMS_table.csv": [],
AvailCap;
# Loading Data to AMPL
read table CUST;
read table DC;
read table MODE;
read table CROP;
read table TECH;
read table WATER;
read table WEEKP;
read table WEEKH;
read table WEEKS;
read table WEEK;

read table LOC;
```

```
read table ZON;
read table YDIST;
read table DEM;
read table CTZC;
read table CTZD;
read table CTDC;
read table CTECH;
read table COL;
read table PARAMS;
read table WEEK1;
read table WEEK2;
read table WEEK3;
read table WEEK4;
read table WEEK5;
read table WEEK6;
option solver cplexamp;
option csvdisplay_header 0;
problem Master:
	#Objective function
	first_stage,
##Constraints
	TechInvL,
```

```
	Tot_land,
	Tot_land_Loc,
	TechType,
	M_Prod,
	Min_Prod,
	 # Cuts
	Cut_Defn,
	##First stage dv's,
	VPlant,
	VB,
	theta;

table params_scen IN "ODBC" "DSN=test" "SQL=SELECT * FROM D:\Dropbox\PhD_Dissertation\AMPLcml\ID\AMPL_RUN\params_scen_BD.csv": [], cur_scen, nCUT;
table yield_sub_table IN "ODBC" "DSN=test" "SQL=SELECT * FROM D:\Dropbox\PhD_Dissertation\AMPLcml\ID\AMPL_RUN\yield_sub_BD.csv": [SCEN,CUTS,WEEKP,CROP,ZON,TECH],yield_sub;
table env_rec_sub_table IN "ODBC" "DSN=test" "SQL=SELECT * FROM D:\Dropbox\PhD_Dissertation\AMPLcml\ID\AMPL_RUN\rain_scen_BD.csv":[WEEKP,WEEKH,ZON],LRainRec;
```

```
table prob_scen_table IN "ODBC" "DSN=test" "SQL=SELECT * FROM D:\Dropbox\PhD_Dissertation\AMPLcml\ID\AMPL_RUN\prob_scen_BD.csv": SCEN <- [SCEN], Prob;
read table prob_scen_table;
read table params_scen;
read table env_rec_sub_table;
read table yield_sub_table;
table micro_harvest_dual_table IN "ODBC" "DSN=test" "SQL=SELECT * FROM D:\Dropbox\PhD_Dissertation\AMPLcml\ID\DUALS\micro_harvest_dual_BD.csv": [SCEN,CUTS,WEEKH,CROP,ZON,TECH,LOC], Micro_harvest_dual;
table Tot_packaging_dual_table IN "ODBC" "DSN=test" "SQL=SELECT * FROM D:\Dropbox\PhD_Dissertation\AMPLcml\ID\DUALS\tot_packaging_dual_BD.csv": [SCEN,CUTS,WEEKH,CROP,QUAL,ZON], Tot_packaging_dual;
table Shipment_L_dual_table IN "ODBC" "DSN=test" "SQL=SELECT * FROM D:\Dropbox\PhD_Dissertation\AMPLcml\ID\DUALS\shipment_l_dual_BD.csv": [SCEN,CUTS,WEEKH,CROP,QUAL,LOC,ZON],Shipment_L_dual;
table Invent_W_dual_table IN "ODBC" "DSN=test" "SQL=SELECT * FROM D:\Dropbox\PhD_Dissertation\AMPLcml\ID\DUALS\invent_w_dual_BD.csv": [SCEN,CUTS,WEEKH,WEEKS,CROP,QUAL,ZON], Invent_W_dual;
table Water_Allocation_dual_table IN "ODBC" "DSN=test" "SQL=SELECT * FROM D:\Dropbox\PhD_Dissertation\AMPLcml\ID\DUALS\water_allocation_dual_BD.csv": [SCEN,CUTS,WEEKP,WEEKH,ZON], Water_Allocation_dual;
```

table Initial_hold_W_dual_table IN "ODBC" "DSN=test" "SQL=SELECT * FROM D:\Dropbox\PhD_Dissertation\AMPLcml\ID\DUALS\initial_hold_w_dual_BD.csv": [SCEN,CUTS,WEEKH,CROP,QUAL,ZON],Initial_hold_W_dual;

table Invent_DC_dual_table IN "ODBC" "DSN=test" "SQL=SELECT * FROM D:\Dropbox\PhD_Dissertation\AMPLcml\ID\DUALS\invent_dc_dual_BD.csv": [SCEN,CUTS,WEEKH,WEEKS,CROP,QUAL,DC], Invent_DC_dual;

table Initial_hold_DC_dual_table IN "ODBC" "DSN=test" "SQL=SELECT * FROM D:\Dropbox\PhD_Dissertation\AMPLcml\ID\DUALS\initial_hold_dc_dual_BD.csv": [SCEN,CUTS,WEEKH,CROP,QUAL,DC],Initial_hold_DC_dual;

table Max_Demand_dual_table IN "ODBC" "DSN=test" "SQL=SELECT * FROM D:\Dropbox\PhD_Dissertation\AMPLcml\ID\DUALS\max_dem_dual_BD.csv": [SCEN,CUTS,CROPS,CUST,WEEKS], Max_Demand_dual;

table Min_Demand_dual_table IN "ODBC" "DSN=test" "SQL=SELECT * FROM D:\Dropbox\PhD_Dissertation\AMPLcml\ID\DUALS\min_dem_dual_BD.csv": [SCEN,CUTS,CROPS,CUST,WEEKS], Min_Demand_dual;

table Ware_Z_Cap_dual_table IN "ODBC" "DSN=test" "SQL=SELECT * FROM D:\Dropbox\PhD_Dissertation\AMPLcml\ID\DUALS\ware_z_cap_dual_BD.csv": [SCEN,CUTS,ZON,WEEKH], Ware_Z_Cap_dual;

table Ship_const_Z_dual_table IN "ODBC" "DSN=test" "SQL=SELECT * FROM D:\Dropbox\PhD_Dissertation\AMPLcml\ID\DUALS\ship_const_z_dual_BD.csv": [SCEN,CUTS,CROP,QUAL,ZON,DC], Ship_const_Z_dual;

```
table Ship_const_W_dual_table IN "ODBC" "DSN=test" "SQL=SELECT * FROM D:\Dropbox\PhD_Dissertation\AMPLcml\ID\DUALS\ship_const_w_dual_BD.csv":
[SCEN,CUTS,CROP,QUAL, ZON,CUST,MOD], Ship_const_W_dual;
table Ship_const_DC_dual_table IN "ODBC" "DSN=test" "SQL=SELECT * FROM D:\Dropbox\PhD_Dissertation\AMPLcml\ID\DUALS\ship_const_dc_dual_BD.csv":
[SCEN,CUTS,CROP,QUAL, DC,CUST,MOD], Ship_const_DC_dual;
read table Initial_hold_W_dual_table;
read table Ship_const_W_dual_table;
read table Ship_const_DC_dual_table;
read table Water_Allocation_dual_table;
read table Initial_hold_DC_dual_table;
read table Max_Demand_dual_table;
read table micro_harvest_dual_table;
read table Shipment_L_dual_table ;
read table Tot_packaging_dual_table;
read table Invent_W_dual_table;
read table Invent_DC_dual_table;
read table Min_Demand_dual_table;
read table Ware_Z_Cap_dual_table;
read table Ship_const_Z_dual_table;
display maxl;
solve Master;
```

```
let{p in WEEKP,j in CROP,l in LOC,u in TECH} VPlant_Output[nCUT,p,j,l,u] := VPlant[p,j,l,u];
let{j in CROP, l in LOC, u in TECH} VB_Output[nCUT,j,l,u] := VB[j,l,u];
let{s in SCEN} VTheta_Output[nCUT,s] := theta[s];
csvdisplay solve_result >> D:\Dropbox\PhD_Dissertation\ AMPLcml\ ID\AMPL_RUN\ master_solve_result_BD.csv;
csvdisplay VPlant >> D: \ Dropbox\ PhD_Dissertation\ AMPLcml\ ID\ AMPL_RUN\ plant_master_BD.csv;
csvdisplay VTheta_Output >> D:\Dropbox\PhD_Dissertation\AMPLcml\ID\AMPL_RUN\ theta_values_BD.csv;
csvdisplay VPlant_Output >> D:\Dropbox\PhD_Dissertation\AMPLcml\ID\AMPL_RUN\ plant_master_plot_BD.csv;
csvdisplay VB_Output >> D:\Dropbox\PhD_Dissertation\AMPLcml\ID\AMPL_RUN\ b_BD.csv;
csvdisplay first_stage >> D:\Dropbox\PhD_Dissertation\AMPLcml\ID\AMPL_RUN\ first_stage_value_BD.csv;
display first_stage;
# Selection of the Model File
reset;
suffix dunbdd OUT;
option display1col 0;
option eexit -10000;
```

```
option solver cplexamp;
model D:\Dropbox\PhD_Dissertation\AMPLcml\ID\Stochastic\BD_V19\ID.mod;
table CUST IN "ODBC" "DSN=test" "SQL=SELECT * FROM D:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\ CUST_table.csv": CUST <-[CUST], Qmin,LT;
table TECH IN "ODBC" "DSN=test" "SQL=SELECT * FROM D:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\TECH.csv": TECH <- [TECH];
table CROP IN "ODBC" "DSN=test" "SQL=SELECT * FROM D:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\CROP_table.csv": CROP <-[CROP], minl,maxl,TraF,Cplant,Ccase,SL;
table LOC IN "ODBC" "DSN=test" "SQL=SELECT * FROM D:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\LOC_table.csv": LOC <-[LOC], cidloc, LAvail;
table ZON IN "ODBC" "DSN=test" "SQL=SELECT * FROM D:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\ZON_table.csv": ZON <-[ZON], cidzone, CTLZ,PZcap,WZ_Cap,Cwater,Cw;
table DC IN "ODBC" "DSN=test" "SQL=SELECT * FROM D:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\DC_table.csv": DC <-[DC],Cd;
table MODE IN "ODBC" "DSN=test" "SQL=SELECT * FROM D:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\MODE_table.csv": MOD <-[MODE];
```

table WEEKP IN "ODBC" "DSN=test" "SQL=SELECT * FROM D:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\WEEKP.csv": WEEKP <- [WEEKP];

table WEEKH IN "ODBC" "DSN=test" "SQL=SELECT * FROM D:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\WEEKH.csv": WEEKH <- [WEEKH];

table WEEKS IN "ODBC" "DSN=test" "SQL=SELECT * FROM D:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\WEEKS.csv": WEEKS <- [WEEKS];

table WEEK IN "ODBC" "DSN=test" "SQL=SELECT * FROM D:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\WEEK.csv": WEEK <- [WEEK];

table WEEK1 IN "ODBC" "DSN=test" "SQL=SELECT * FROM D:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\WEEK1.csv": WEEK1 <- [WEEKH,WEEKS];

table WEEK2 IN "ODBC" "DSN=test" "SQL=SELECT * FROM D:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\WEEK2.csv": WEEK2 <- [WEEKP, WEEKH,CROP,ZON,TECH];

table WEEK3 IN "ODBC" "DSN=test" "SQL=SELECT * FROM D:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\WEEK3.csv": WEEK3 <- [WEEKP, WEEKH,CROP,ZON];

table WEEK4 IN "ODBC" "DSN=test" "SQL=SELECT * FROM D:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\WEEK4.csv": WEEK4 <- [WEEKP, WEEKH,CROP];

table WEEK5 IN "ODBC" "DSN=test" "SQL=SELECT * FROM D:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\WEEK5.csv": WEEK5 <- [WEEKP, WEEKH,ZON];

table WEEK6 IN "ODBC" "DSN=test" "SQL=SELECT * FROM D:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\WEEK6.csv": WEEK6 <- [WEEKP, WEEKH];

table YDIST IN "ODBC" "DSN=test" "SQL=SELECT * FROM D:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\YDist_table.csv": [WEEKP,WEEKH,CROP,ZON,TECH], YDist;

table DEM IN "ODBC" "DSN=test" "SQL=SELECT * FROM D:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\DEM_table.csv": [WEEKS,CROP,CUST], MaxDem, MinDem;

table CTZC IN "ODBC" "DSN=test" "SQL=SELECT * FROM D:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\CTZC_table.csv": [ZON,CUST,MOD], CTZC,TimeZC;

table CTZD IN "ODBC" "DSN=test" "SQL=SELECT * FROM D:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\CTZD_table.csv": [ZON,DC], CTZD,TimeZD;

```
table CTDC IN "ODBC" "DSN=test" "SQL=SELECT * FROM
D:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\CTDC_table.csv":
[DC,CUST,MOD], CTDC,TimeDC;
table CTECH IN "ODBC" "DSN=test" "SQL=SELECT * FROM
D:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\CTECH_table.csv":
[TECH,ZON], Ctech,Coper;
table COL IN "ODBC" "DSN=test" "SQL=SELECT * FROM
D:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\COL_table.csv":
[WEEKH,CROP,QUAL], COL;
table WATER IN "ODBC" "DSN=test" "SQL=SELECT * FROM
D:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\WATER_REQ.csv":[CROP,TEC
H], MinWReq;
#table ENVREC IN "ODBC" "DSN=test" "SQL=SELECT * FROM
D:\Dropbox\PhD_Dissertation\AMPLcml\ID\Datasets\ENVREC_table.csv":[WEEKP,W
EEKH,ZON],LRainRec;
# Loading Data to AMPL
read table CUST;
read table DC;
read table MODE;
read table CROP;
read table TECH;
read table WATER;
```

```
read table WEEKP;
read table WEEKH;
read table WEEKS;
read table WEEK;
read table WEEK1;
read table WEEK2;
read table WEEK3;
read table WEEK4;
read table WEEK5;
read table WEEK6;
read table LOC;
read table ZON;
read table YDIST;
read table DEM;
read table CTZC;
read table CTZD;
read table CTDC;
read table CTECH;
read table COL;
option solver cplexamp;
option csvdisplay_header 0;
# Definition of Subproblem #
```

```
problem Sub:
        #Objective funtion
        second_stage,
        ##Constraints
        Micro_harvest,
    Tot_packaging,
  Shipment_L,
  Ship_const_Z,
  Ship_const_W,
  Ship_const_DC,
 Max_Demand,
 Min_Demand,
 Water_Allocation,
       Ware_Z_Cap,
       Add_War,
        ##Second stage dv's
        Initial_hold_W,
        Invent_W,
        Initial_hold_DC,
        Invent_DC,
        Inv_Const_W ,
        Inv_Const_DC,
```

```
        VSZD,
    VSDC,
    VSLZ,
    VPACK,
    VMicroHarv,
    VInvw,
    VInvd,
    VWatAll
        ;
table prob_scen_table "ODBC" "DSN=test" "SQL=SELECT * FROM D:\Dropbox\PhD_Dissertation\AMPLcml\ID\AMPL_RUN\prob_scen_BD.csv": SCEN <- [SCEN], Prob;
table yields_scen "ODBC" "DSN=test" "SQL=SELECT * FROM D:\Dropbox\PhD_Dissertation\AMPLcml\ID\AMPL_RUN\yields_scen_BD.csv": [WEEKP,CROP,ZON,TECH], yield;
table prices_scen "ODBC" "DSN=test" "SQL=SELECT * FROM D:\Dropbox\PhD_Dissertation\AMPLcml\ID\AMPL_RUN\prices_scen_BD.csv": [WEEKS,CROP,CUST], price;
table params_scen "ODBC" "DSN=test" "SQL=SELECT * FROM D:\Dropbox\PhD_Dissertation\AMPLcml\ID\AMPL_RUN\params_scen_BD.csv": [], cur_scen, nCUT;
```

```
table plant_master "ODBC" "DSN=test" "SQL=SELECT * FROM
D:\Dropbox\PhD_Dissertation\AMPLcml\ID\AMPL_RUN\plant_master_BD.csv":
[WEEKP,CROP,LOC,TECH], plant_master;
table env_rec_sub_table "ODBC" "DSN=test" "SQL=SELECT * FROM
D:\Dropbox\PhD_Dissertation\AMPLcml\ID\AMPL_RUN\rain_scen_BD.csv":[WEEKP
,WEEKH,ZON],LRainRec;
read table env_rec_sub_table;
read table prob_scen_table;
read table params_scen;
read table yields_scen;
read table plant_master;
read table prices_scen;
solve Sub;
csvdisplay solve_result;
csvdisplay solve_result >> D:\Dropbox\PhD_Dissertation\AMPLcml\ID\AMPL_RUN\
solve_result_BD.csv;
let value :=
sum{z in ZON, h in WEEKH} Ware_Z_Cap[z,h].dual*WZ_Cap[z]
+ sum{j in CROP,i in CUST,t in WEEKS} Max_Demand[j,i,t].dual*MaxDem[t,j,i]
+ sum{j in CROP,i in CUST,t in WEEKS} Min_Demand[j,i,t].dual*-1*MinDem[t,j,i]
+ sum{h in WEEKH, j in CROP, q in QUAL,l in LOC, z in ZON:
cidloc[l]==cidzone[z]}Shipment_L[h,j,q,l,z].dual
```

```
+ sum{j in CROP,q in QUAL,z in ZON, d in DC} Ship_const_Z[j,q,z,d].dual
+ sum {j in CROP,q in QUAL,z in ZON, i in CUST, m in MOD}
Ship_const_W[j,q,z,i,m].dual
+ sum {j in CROP,q in QUAL,d in DC, i in CUST, m in MOD}
Ship_const_DC[j,q,d,i,m].dual
+ sum{h in WEEKH,j in CROP,q in QUAL,z in ZON} Initial_hold_W[h,j,q,z].dual
+ sum{(h,t) in WEEK1,j in CROP,q in QUAL,z in ZON: t>h}Invent_W[h,t,j,q,z].dual
+ sum{h in WEEKH,j in CROP,q in QUAL,d in DC}Initial_hold_DC[h,j,q,d].dual
+ sum{(h,t) in WEEK1,j in CROP,q in QUAL,d in DC: t>h}Invent_DC[h,t,j,q,d].dual
+ sum{h in WEEKH,j in CROP,q in QUAL,z in ZON} Tot_packaging[h,j,q,z].dual
+ sum{(p,h,z) in WEEK5} Water_Allocation[p,h,z].dual*-1*lrainrec[p,h,z]
- sum{(p,h) in WEEK6, j in CROP, z in ZON, u in TECH, l in LOC:cidloc[l]==cidzone[z]}
Micro_harvest[h,j,z,u,l].dual *(plant_master[p,j,l,u]*YDist[p,h,j,z,u]*yield[p,j,z,u])
- sum{(p,h,z) in WEEK5} Water_Allocation[p,h,z].dual*sum{l in LOC, u in TECH, j in
CROP: cidloc[l]==cidzone[z]} ( MinWReq[j,u] * plant_master[p,j,l,u])
  ;
display value;
let {h in WEEKH,j in CROP,q in QUAL,d in DC}
Initial_hold_DC_dual[cur_scen,nCUT,h,j,q,d] := Initial_hold_DC[h,j,q,d].dual;
let {(h,t) in WEEK1,j in CROP,q in QUAL,d in DC:
t>h}Invent_DC_dual[cur_scen,nCUT,h,t,j,q,d]:=  Invent_DC[h,t,j,q,d].dual;
```

```
let {j in CROP,i in CUST,t in WEEKS} Max_Demand_dual[cur_scen,nCUT,j,i,t] :=
Max_Demand[j,i,t].dual;
let {j in CROP,i in CUST,t in WEEKS} Min_Demand_dual[cur_scen,nCUT,j,i,t] :=
Min_Demand[j,i,t].dual;
let {z in ZON, h in WEEKH} Ware_Z_Cap_dual[cur_scen,nCUT,z,h] :=
Ware_Z_Cap[z,h].dual;
let {j in CROP,q in QUAL,z in ZON, d in DC} Ship_const_Z_dual[cur_scen,nCUT,j,q,z,d]
:= Ship_const_Z[j,q,z,d].dual;
let {j in CROP,q in QUAL,z in ZON, i in CUST, m in MOD}
Ship_const_W_dual[cur_scen,nCUT,j,q,z,i,m] := Ship_const_W[j,q,z,i,m].dual;
let {j in CROP,q in QUAL,d in DC, i in CUST, m in MOD}
Ship_const_DC_dual[cur_scen,nCUT,j,q,d,i,m] := Ship_const_DC[j,q,d,i,m].dual;
let {(p,h,z) in WEEK5} Water_Allocation_dual[cur_scen,nCUT,p,h,z] :=
Water_Allocation[p,h,z].dual;
let {h in WEEKH, j in CROP, z in ZON, u in TECH, l in LOC:cidloc[l]==cidzone[z]}
Micro_harvest_dual[cur_scen,nCUT,h,j,z,u,l] := Micro_harvest[h,j,z,u,l].dual;
let {h in WEEKH,j in CROP,q in QUAL,z in ZON}
Tot_packaging_dual[cur_scen,nCUT,h,j,q,z] := Tot_packaging[h,j,q,z].dual;
let {(h,t) in WEEK1,j in CROP,q in QUAL,z in ZON:
t>h}Invent_W_dual[cur_scen,nCUT,h,t,j,q,z]:= Invent_W[h,t,j,q,z].dual;
let {h in WEEKH,j in CROP,q in QUAL,l in LOC,z in ZON: cidloc[l]==cidzone[z] }
Shipment_L_dual[cur_scen,nCUT,h,j,q,l,z] := Shipment_L[h,j,q,l,z].dual;
```

```
let {h in WEEKH,j in CROP,q in QUAL,z in ZON}
Initial_hold_W_dual[cur_scen,nCUT,h,j,q,z] := Initial_hold_W[h,j,q,z].dual;
csvdisplay Initial_hold_DC_dual >>
D:\Dropbox\PhD_Dissertation\AMPLcml\ID\DUALS\ initial_hold_dc_dual_BD.csv;
csvdisplay Max_Demand_dual >> D:\Dropbox\PhD_Dissertation\AMPLcml\ID\DUALS\
max_dem_dual_BD.csv;
csvdisplay Min_Demand_dual >> D:\Dropbox\PhD_Dissertation\AMPLcml\ID\DUALS\
min_dem_dual_BD.csv;
csvdisplay Ware_Z_Cap_dual >> D:\Dropbox\PhD_Dissertation\AMPLcml\ID\DUALS\
ware_z_cap_dual_BD.csv;
csvdisplay Ship_const_Z_dual >> D:\Dropbox\PhD_Dissertation\AMPLcml\ID\DUALS\
ship_const_z_dual_BD.csv;
csvdisplay Ship_const_W_dual >>
D:\Dropbox\PhD_Dissertation\AMPLcml\ID\DUALS\ ship_const_w_dual_BD.csv;
csvdisplay Ship_const_DC_dual >>
D:\Dropbox\PhD_Dissertation\AMPLcml\ID\DUALS\  ship_const_dc_dual_BD.csv;
csvdisplay Water_Allocation_dual>>
D:\Dropbox\PhD_Dissertation\AMPLcml\ID\DUALS\ water_allocation_dual_BD.csv;
csvdisplay Micro_harvest_dual >> D:\Dropbox\PhD_Dissertation\AMPLcml\ID\DUALS\
micro_harvest_dual_BD.csv;
csvdisplay Tot_packaging_dual >>
D:\Dropbox\PhD_Dissertation\AMPLcml\ID\DUALS\ tot_packaging_dual_BD.csv;
```

```
csvdisplay Shipment_L_dual >> D:\Dropbox\PhD_Dissertation\AMPLcml\ID\DUALS\
shipment_l_dual_BD.csv;
csvdisplay Invent_W_dual >> D:\Dropbox\PhD_Dissertation\AMPLcml\ID\DUALS\
invent_w_dual_BD.csv;
csvdisplay Invent_DC_dual >> D:\Dropbox\PhD_Dissertation\AMPLcml\ID\DUALS\
invent_dc_dual_BD.csv;
csvdisplay Initial_hold_W_dual >>
D:\Dropbox\PhD_Dissertation\AMPLcml\ID\DUALS\ initial_hold_w_dual_BD.csv;
csvdisplay VMicroHarv >> D:\Dropbox\PhD_Dissertation\AMPLcml\ID\OUTPUT\
vmicroharv_BD.csv;
csvdisplay VSDC >>
D:\Dropbox\PhD_Dissertation\AMPLcml\ID\OUTPUT\vsdc_BD.csv;
csvdisplay VInvw >> D:\Dropbox\PhD_Dissertation\AMPLcml\ID\OUTPUT\
vinvw_stoch_BD.csv;
csvdisplay VInvd >> D:\Dropbox\PhD_Dissertation\AMPLcml\ID\OUTPUT\
vinvd_stoch_BD.csv;
#####
#####
# R – Script
#
####
start_tables_BD <- function(){
```

```
Water_Allocation_dual_table  <- matrix(0,ncol = 6);Water_Allocation_dual_table [1,] <- c('SCEN','CUTS','WEEKP','WEEKH','ZON','Water_Allocation_dual');
write.table(Water_Allocation_dual_table ,file = '../../../AMPLcml/ID/DUALS/water_allocation_dual_BD.csv',sep=',', row.names=FALSE,col.names = FALSE)
micro_harvest_dual_table <- matrix(0,ncol = 8);micro_harvest_dual_table[1,] <- c('SCEN','CUTS','WEEKH','CROP','ZON','TECH','LOC','Micro_harvest_dual');
write.table(micro_harvest_dual_table,file = '../../../AMPLcml/ID/DUALS/micro_harvest_dual_BD.csv',sep=',', row.names=FALSE,col.names = FALSE)
 Tot_packaging_dual_table <- matrix(0,ncol = 7);Tot_packaging_dual_table[1,] <- c('SCEN','CUTS','WEEKH','CROP','QUAL','ZON','Tot_packaging_dual');
  write.table(Tot_packaging_dual_table,file = '../../../AMPLcml/ID/DUALS/tot_packaging_dual_BD.csv',sep=',', row.names=FALSE,col.names = FALSE)

  Shipment_L_dual_table <- matrix(0,ncol = 8);Shipment_L_dual_table[1,] <- c('SCEN','CUTS','WEEKH','CROP','QUAL','LOC','ZON','Shipment_L_dual');
  write.table(Shipment_L_dual_table,file = '../../../AMPLcml/ID/DUALS/shipment_l_dual_BD.csv',sep=',', row.names=FALSE,col.names = FALSE)
Max_Demand_dual_table <- matrix(0,ncol = 6);
Max_Demand_dual_table[1,] <- c('SCEN','CUTS','CROPS','CUST','WEEKS', 'Max_Demand_dual');
```

```
write.table(Max_Demand_dual_table,file = '../../../AMPLcml/ID/DUALS/
max_dem_dual_BD.csv',sep=',', row.names=FALSE,col.names = FALSE)
Min_Demand_dual_table <- matrix(0,ncol = 6);Min_Demand_dual_table[1,] <-
c('SCEN','CUTS','CROPS','CUST','WEEKS','Min_Demand_dual');
 write.table(Min_Demand_dual_table,file =
'../../../AMPLcml/ID/DUALS/min_dem_dual_BD.csv',sep=',',
row.names=FALSE,col.names = FALSE)
Ware_Z_Cap_dual_table <- matrix(0,ncol = 5);Ware_Z_Cap_dual_table[1,] <-
c('SCEN','CUTS','ZON','WEEKH','Ware_Z_Cap_dual');
 write.table(Ware_Z_Cap_dual_table,file =
'../../../AMPLcml/ID/DUALS/ware_z_cap_dual_BD.csv',sep=',',
row.names=FALSE,col.names = FALSE)
Initial_hold_W_dual_table <- matrix(0,ncol = 7);Initial_hold_W_dual_table[1,] <-
c('SCEN','CUTS','WEEKH','CROP','QUAL','ZON','Initial_hold_W_dual');
write.table(Initial_hold_W_dual_table,file =
'../../../AMPLcml/ID/DUALS/initial_hold_w_dual_BD.csv',sep=',',
row.names=FALSE,col.names = FALSE)
Initial_hold_DC_dual_table <- matrix(0,ncol = 7);Initial_hold_DC_dual_table[1,] <-
c('SCEN','CUTS','WEEKH','CROP','QUAL','DC','Initial_hold_DC_dual');
write.table(Initial_hold_DC_dual_table,file =
'../../../AMPLcml/ID/DUALS/initial_hold_dc_dual_BD.csv',sep=',',
row.names=FALSE,col.names = FALSE)
```

```
Invent_W_dual_table <- matrix(0,ncol = 8);Invent_W_dual_table[1,] <-
c('SCEN','CUTS','WEEKH','WEEKS','CROP','QUAL','ZON','Invent_W_dual');
write.table(Invent_W_dual_table,file =
'../../../AMPLcml/ID/DUALS/invent_w_dual_BD.csv',sep=',',
row.names=FALSE,col.names = FALSE)
Invent_DC_dual_table <- matrix(0,ncol = 8);Invent_DC_dual_table[1,] <-
c('SCEN','CUTS','WEEKH','WEEKS','CROP','QUAL','DC','Invent_DC_dual');
write.table(Invent_DC_dual_table,file = '../../../AMPLcml/ID/DUALS/
invent_dc_dual_BD.csv',sep=',', row.names=FALSE,col.names = FALSE)
Ship_const_Z_dual_table <- matrix(0,ncol = 7);Ship_const_Z_dual_table[1,] <-
c('SCEN','CUTS','CROP','QUAL','ZON','DC','Ship_const_Z_dual');
write.table(Ship_const_Z_dual_table,file = '../../../AMPLcml/ID/DUALS/
ship_const_z_dual_BD.csv',sep=',', row.names=FALSE,col.names = FALSE)
Ship_const_W_dual_table <- matrix(0,ncol = 8);Ship_const_W_dual_table[1,] <-
c('SCEN','CUTS','CROP','QUAL','ZON','CUST','MOD','Ship_const_W_dual');
write.table(Ship_const_W_dual_table,file =
'../../../AMPLcml/ID/DUALS/ship_const_w_dual_BD.csv',sep=',',
row.names=FALSE,col.names = FALSE)
Ship_const_DC_dual_table <- matrix(0,ncol = 8);Ship_const_DC_dual_table[1,] <-
c('SCEN','CUTS','CROP','QUAL','DC','CUST','MOD','Ship_const_DC_dual');
```

```
write.table(Ship_const_DC_dual_table,file =
'../../../AMPLcml/ID/DUALS/ship_const_dc_dual_BD.csv',sep=',',
row.names=FALSE,col.names = FALSE)
write.csv(NULL,file='../../../AMPLcml/ID/AMPL_RUN/theta_values_BD.csv',row.name
s = FALSE)
t <- data.frame(1); names(t) <- c('Master')
write.csv(t,file='../../../AMPLcml/ID/AMPL_RUN/first_stage_value_BD.csv',row.names
= FALSE);
write.csv(NULL,file='../../../AMPLcml/ID/AMPL_RUN/plant_master_plot_BD.csv',row.
names = FALSE)
write.csv(NULL,file='../../../AMPLcml/ID/OUTPUT/vmicroharv_BD.csv',row.names =
FALSE)
write.csv(NULL,file='../../../AMPLcml/ID/OUTPUT/vslz_BD.csv',row.names = FALSE)
write.csv(NULL,file='../../../AMPLcml/ID/OUTPUT/vsdc_BD.csv',row.names = FALSE)
write.csv(NULL,file='../../../AMPLcml/ID/OUTPUT/vszc_BD.csv',row.names = FALSE)
write.csv(NULL,file='../../../AMPLcml/ID/OUTPUT/vinvw_stoch_BD.csv',row.names =
FALSE)
write.csv(NULL,file='../../../AMPLcml/ID/OUTPUT/vinvd_stoch_BD.csv',row.names =
FALSE)
write.csv(NULL,file='../../../AMPLcml/ID/OUTPUT/plant_deter_BD.csv',row.names =
FALSE)
```

```
write.csv(NULL,file='../../../AMPLcml/ID/AMPL_RUN/plant_master_BD.csv',row.name
s = FALSE)
write.csv(NULL,file='../../../AMPLcml/ID/AMPL_RUN/b_BD.csv',row.names = FALSE)
write.csv(NULL,file='../../../AMPLcml/ID/AMPL_RUN/plant_plot_BD.csv',row.names =
FALSE)
write.csv(NULL,file='../../../AMPLcml/ID/AMPL_RUN/rain_scen_BD.csv',row.names =
FALSE)
write.csv(NULL,file='../../../AMPLcml/ID/AMPL_RUN/yields_scen_BD.csv',row.names
= FALSE)
write.csv(NULL,file='../../../AMPLcml/ID/AMPL_RUN/prices_scen_BD.csv',row.names
= FALSE)
write.csv(NULL,file='../../../AMPLcml/ID/OUTPUT/MicroHarv_deter_BD.csv',row.nam
es = FALSE)
write.csv(NULL,file='../../../AMPLcml/ID/OUTPUT/szc_deter_BD.csv',row.names =
FALSE)
write.csv(NULL,file = '../../../AMPLcml/ID/AMPL_RUN/yield_sub_BD.csv')
}
#####

#####
rm(list=ls())
source('Crear_Sets_Tables.R')
```

```
source('Run_AMPL_SetUP.R')
prices <- data.frame(read.csv(file = '../../../AMPLcml/ID/Datasets/PRICES_table.csv'));
names(prices)[ncol(prices)] <- c('price')  # reads in sampled data that has already been constructed.  500 samples were initially created
yields_sample <- read.csv(file = '../../../AMPLcml/ID/Datasets/Yields_sampled.csv')
rain_sample <- read.csv(file='../../../AMPLcml/ID/Datasets/PRCP_scen_table.csv')
ZON <- data.frame(read.csv(file = '../../../AMPLcml/ID/Datasets/ZON_table.csv'))[,1]
TECH <- data.frame(read.csv(file= '../../../AMPLcml/ID/Datasets/TECH.csv'))[,1]
CROP <- data.frame(read.csv(file= '../../../AMPLcml/ID/Datasets/CROP_table.csv'))[,1]
#Load_Sets()
run_libraries()
start_tables_BD()
#######################
#
version <- 'BD_V19'
cur_path <- dirname(dirname(dirname(getwd()))); path_to_directory <- cur_path
cur_path <- gsub('/','\\\\',cur_path)
cur_path <- paste(cur_path,'AMPLcml','ID','Stochastic',version,sep='\\');
path_to_model <- cur_path
cur_path <- paste('ampl',cur_path)
#########################
update_paths_BD()
```

```
## NOTE: Remember that each scenario contain randomly generated prices for all markets
## the same goes for yields. Each scenario represents the expected yields for each zone
according to different temperatures
## It should also be noted that each scenario is represented by a vector of 52 weekly values.
sample_size <- 20
random_sample_scen <- sample(length(unique(yields_sample$ITER)),sample_size)
yields_sample_subset <-
yields_sample[yields_sample$ITER%in%random_sample_scen,]
prices_sample <- prices[prices$SCEN%in%random_sample_scen,]
prices_sample <- prices_sample[order(prices_sample$WEEKS,prices_sample$SCEN,
prices_sample$CROP),]
rain_sample_subset <- rain_sample[rain_sample$SCEN%in%random_sample_scen,]
scenarios <- data.frame(unique(yields_sample_subset$ITER))
names(scenarios) <- c('SCEN')  # simple number for each scenario
scenarios$Prob <- 1/sample_size
write.csv(scenarios,file='../../../AMPLcml/ID/AMPL_RUN/prob_scen_BD.csv',
row.names = FALSE)
yield_sub <- NULL
prob_scen <- NULL
master_values <- NULL
nCut <- 0
tot_iter <- 120
```

```
for (i in 1:tot_iter) {
  if(i>120) {  break }
  print(paste('Iteration',i,sep=':'))
  nCut <- i
  print(paste('CUT',nCut,sep = ' '))
  for(j in 1:length(random_sample_scen))  {
          print(paste('CUT',nCut,sep = ' '))
            s <- random_sample_scen[j]
           test_solved <-'not_solved';
          force_write_csv(NULL,'../../../AMPLcml/ID/AMPL_RUN/solve_result_BD.csv')
          print(paste('SCENARIO',s,':',j/nrow(random_sample_scen)*100,'%',sep=' '))
          if(nCut==1)
          {
            plant_master <- data.frame(read.csv('../../../AMPLcml/ID/AMPL_RUN/
          plant_master_iter_0_BD.csv',header = TRUE))
            force_write_csv(plant_master,'../../../AMPLcml/ID/AMPL_RUN/
          plant_master_BD.csv')
          }
          plant_master_und <- read.csv('../../../AMPLcml/ID/AMPL_RUN/
          plant_master_BD.csv')
          plant_master_und <- plant_master_und[!(duplicated(plant_master_und)),]
          Plant <- read.csv('../../../AMPLcml/ID/AMPL_RUN/
```

```
plant_master_plot_BD.csv',header = FALSE)
colnames(Plant) <- c('CUT','WEEKP','CROP','LOC','TECH','plant_master')
Plant <- Plant[!(duplicated(Plant[,1:(ncol(Plant)-1)])),]
Plant <- Plant[Plant$CUT==max(unique(Plant$CUT)),]
Plant <- Plant[,-1]
plant_master_und <- Plant
force_write_csv(plant_master_und,'../../../AMPLcml/ID/
AMPL_RUN/plant_master_BD.csv')
prices_scen <- prices_sample[prices_sample$SCEN==s,-1]
yields_scen <- yields_sample_subset[yields_sample_subset$ITER==s,
names(yields_sample_subset)%in%c('WEEKP','WEEKH','CROP',
'ZON','TECH','Yield')            ]
names(yields_scen)[ncol(yields_scen)] <- c('yield')
yields_scen.agg <- aggregate(yield~WEEKP+CROP+ZON+TECH+
yield,yields_scen,sum)
rain_scen <- rain_sample_subset[rain_sample_subset$SCEN==s,
names(rain_sample_subset)%in%c('WEEKP','WEEKH','ZON','LRainRec')]
params <- data.frame(t(c(s,nCut))); names(params) <- c('cur_scen','nCUT')
force_write_csv(params,'../../../AMPLcml/ID/AMPL_RUN/
params_scen_BD.csv')
force_write_csv(prices_scen,'../../../AMPLcml/ID/AMPL_RUN/
prices_scen_BD.csv')
```

```
        force_write_csv(yields_scen.agg,'../../../AMPLcml/ID/AMPL_RUN/
         yields_scen_BD.csv')
         force_write_csv(rain_scen,'../../../AMPLcml/ID/AMPL_RUN/rain_scen_BD.csv')
         while(test_solved != 'solved')  ## solves ampl until it is able to correctly run
         {
           path_submodel <- paste(cur_path,'\\sub_model.mod',sep='')
           print(shell(path_submodel))
           test_solved <- read.csv('../../../AMPLcml/ID/AMPL_RUN/
         solve_result_BD.csv',header = FALSE,stringsAsFactors = FALSE)
          test_solved <- substr(test_solved,nchar(test_solved)-5,nchar(test_solved))
          print(test_solved)
          force_write_csv(NULL,'../../../AMPLcml/ID/AMPL_RUN/
         solve_result_BD.csv')
          }
         ## for master problem
         names(yields_scen.agg)[ncol(yields_scen.agg)] <- c('yield_sub')
         yields_scen.agg$SCEN <-  s
         yields_scen.agg$CUTS <- nCut
    yield_sub <- rbind(yield_sub,yields_scen.agg[, c('SCEN','CUTS','WEEKP','CROP',
'ZON','TECH','yield_sub')])
    if(i>1)
    {
```

```
        write.csv(NULL,file='../../../AMPLcml/ID/OUTPUT/
        vmicroharv_BD.csv',row.names = FALSE)
         write.csv(NULL,file='../../../AMPLcml/ID/OUTPUT/
        vinvw_stoch_BD.csv',row.names = FALSE)
      write.csv(NULL,file='../../../AMPLcml/ID/OUTPUT/
        vinvd_stoch_BD.csv',row.names = FALSE)
        write.csv(NULL,file='../../../AMPLcml/ID/OUTPUT/
        vsdc_BD.csv',row.names = FALSE)
    } }
  order_duals_BD()
  yield_sub <- yield_sub[!(duplicated(yield_sub[,1:(ncol(yield_sub))])),]
   force_write_csv(yield_sub,file = '../../../AMPLcml/ID/AMPL_RUN/yield_sub_BD.csv')
## stores the yields per cut and scenario..used in the cuts constraints
plant_master_table  <- matrix(0,ncol = 5)
plant_master_table[1,] <- c('WEEKP','CROP','LOC','TECH','plant_master');
write.table(plant_master_table ,file = '../../../AMPLcml/ID/AMPL_RUN/
plant_master_BD.csv',sep=',', row.names=FALSE,col.names = FALSE)
print('MASTER')
#### solve master problem
master_test_solved<-'not_solved';
write.csv(NULL,file='../../../AMPLcml/ID/AMPL_RUN/master_solve_result_BD.csv');
# clears the solve display from ampl
```

```
while(master_test_solved != 'solved')  ## solves ampl until it is able to correctly run
{  path_mastermodel <- paste(cur_path,'\\master_model.mod',sep='')
    print(shell(path_mastermodel))
    master_test_solved <- read.csv('../../../AMPLcml/ID/AMPL_RUN/
    master_solve_result_BD.csv',header = FALSE,stringsAsFactors = FALSE)
    master_test_solved <- substr(master_test_solved,
   nchar(master_test_solved)-5,nchar(master_test_solved))
    print(master_test_solved)
    force_write_csv(NULL,'../../../AMPLcml/ID/AMPL_RUN/
    master_solve_result_BD.csv')}
  plant_master_reordered <- data.frame(read.csv( '../../../AMPLcml/ID/AMPL_RUN/
   plant_master_BD.csv'))
  plant_master_reordered <- plant_master_reordered[order(-plant_master_reordered
  $plant_master),]
  plant_master_reordered <- plant_master_reordered[!(duplicated(
   plant_master_reordered)),]
  force_write_csv(plant_master_reordered,'../../../AMPLcml/ID/AMPL_RUN/
   plant_master_BD.csv')



}
```

APPENDIX D

STOCHASTIC DECOMPOSITION FORMULATION

```
#param directory;

#reset;

# Data sets #

set WEEKP ordered;                  # weeks of the planning period

set WEEKH ordered;                  # weeks of the harvesting period

set CROP;                    # Crops for planting

set CUST;                     # Customers

set TECH;                          # Growing technologies

set ZON;                     # Zones to be considered

set LOC;                    # Locations available for planting cropss

set SCEN;

# Second-stage scenarios

set DC;

set MOD;

#####

set WEEK ordered;

set WEEKS ordered;

#Additional necessary sets to facilitate indexing

set WEEK1 within {WEEKH, WEEKS};

set WEEK2 within {WEEKP, WEEKH,CROP,ZON,TECH};

set WEEK3 within {WEEKP, WEEKH,CROP,ZON};

set WEEK4 within {WEEKP, WEEKH,CROP};
```

```
set WEEK5 within {WEEKP, WEEKH,ZON};

set WEEK6 within {WEEKP, WEEKH};

set QUAL:= 2..2;                    # Color characteristic of products

param cur_scen >=0;

set ITER:= 1..cur_scen;

param NbScen >= 0;

########

# Parameter definition #

# Production/Yield-Related

param YDist {WEEK2} >= 0;     # Expected YDist per crop

param yield{SCEN,WEEKP,CROP,ZON,TECH} >= 0;

param lrainrec{WEEKP,WEEKH,ZON} default 0;

param Salv {WEEKP,WEEKH,CROP,ZON,TECH} >=0;  # Expected waste of crops

param MaxDem {WEEKS,CROP,CUST} >= 0;

param MinDem {WEEKS,CROP,CUST} >= 0;

param Qmin {CUST}>=0;                    # Quality demanded by customer i

param COL {WEEKH,CROP,QUAL}>=0;

param Maxi{LOC} >=0;# Maximum amount of hectares to plant in a given period

param minl{CROP} >=0;                    # Minimum amount to plant per crop

param maxl{CROP} >=0;                    # Maximum amount to plant per crop

param Ctech{TECH, ZON} >=0;                 # Cost of technology u for location l

param Coper{TECH, ZON} >= 0;
```

```
param Cplant{CROP} >=0;
param Cwater{ZON}  >=0;                    # Cost of water per location l
param cidloc {LOC} symbolic;
param cidzone{ZON} symbolic;
# Environment-Related
param MinWReq {CROP,TECH} >=0;
param LAvail {LOC} >=0;
# Logistics Related
param SL {CROP} >=0;                 # Shelf life of product k
param LT {CUST} >=0;              # Lead time required by the customer
param Weight {CROP} >=0;                 # Quantity in required of crop j to form a case
of product k
param PZcap {ZON} >=0;              # Capacity of te packaging facility for a time period
param TimeZC {ZON,CUST,MOD} >=0;
param TimeDC {DC,CUST,MOD} >=0;
param TimeZD {ZON,DC} >=0;
param CTZC{ZON,CUST,MOD} >=0;
param CTZD{ZON,DC} >=0;
param CTDC{DC,CUST,MOD} >=0;
param CTLZ{ZON} >=0;
param Cw {ZON} >= 0; # Cost of warehouse
param Cd {DC} >= 0;
```

```
param Ccase{CROP} >=0;                  # Cost for packing case for crop j
param TraF{CROP}  >=0;
param WZ_Cap{ZON} >= 0;
# Market Related
param price {WEEKS,CROP,CUST} >= 0;
# Additional parameters                 # Investment available for the season
param cur_CUT >=0;
# Water available
param AvailCap >=0;                     # Capital available for technology investments
# Helper parameters for outputs
param PZC {CROP,LOC,ZON,TECH} ;
param nCUT >= 0 integer;# Counter of optimality cuts
param fCUT >= 0 integer;
param Type_cut >= 0 integer;
param count default 0;
param theta_k{SCEN} default 0;
param break_check default 0;
param counter_last default 0;
set CUTS:=1..nCUT;
param value default 0;
####### Stochastic parameters and shadow prices
param Prob{SCEN} >=0;
```

```
param yield_sub {SCEN,WEEKP,CROP,ZON,TECH} >= 0;
param plant_master {WEEKP,CROP,LOC,TECH} >= 0;
param LRainRec{SCEN,WEEKP,WEEKH,ZON} >= 0;

param Water_Allocation_dual {1..nCUT, WEEK5};
param Zone_harvest_dual {1..nCUT,WEEK3};
param Micro_harvest_dual{1..nCUT,WEEKH, CROP,z in ZON, TECH,l in LOC:cidloc[l]==cidzone[z]};
param Shipment_L_dual {1..nCUT, WEEKH, CROP,QUAL,l in LOC,z in ZON: cidloc[l]==cidzone[z]};
param Tot_packaging_dual {1..nCUT, WEEKH,CROP, QUAL,ZON};
param Invent_W_dual {1..nCUT,(h,t) in WEEK1,CROP,QUAL,ZON:t>h};
param Initial_hold_W_dual {1..nCUT, WEEKH,CROP,QUAL,ZON};
param Invent_DC_dual {1..nCUT,(h,t) in WEEK1,CROP,QUAL,DC:t>h};
param Initial_hold_DC_dual {1..nCUT, WEEKH,CROP,QUAL,DC};
param Max_Demand_dual {1..nCUT,CROP,CUST,WEEKS};
param Min_Demand_dual {1..nCUT,CROP,CUST,WEEKS};
param Ware_Z_Cap_dual {1..nCUT,ZON,WEEKH};
param Ship_const_Z_dual {1..nCUT,CROP,QUAL,ZON,DC};
param Ship_const_W_dual {1..nCUT,CROP,QUAL,ZON,CUST,MOD};
param Ship_const_DC_dual {1..nCUT,CROP,QUAL,DC,CUST,MOD};
param Max_Profits_dual{1..nCUT};
```

```
param VMicroHarv_Output {1..nCUT, WEEKH, CROP, LOC, TECH};

param VPlant_Output {1..nCUT,WEEKP,CROP,LOC,TECH};

param VTheta_Output {1..nCUT};

########## Variable Definition

#Planting and resource decisions

var VPlant {WEEKP,CROP,LOC,TECH} >=0;

var VMicroHarv {WEEKH,CROP,LOC,TECH} >=0;     # Harvest (pounds)

var VZoneHarv {(p,h,j,z) in WEEK3}  >=0;

var SumPerishable>=0;

var VWatAll {WEEK5} >=0;

var VB {CROP,LOC,TECH} >= 0 binary;

#Logistic related variables

var VPACK {WEEKH,CROP,QUAL,ZON} >=0;

var VSLZ {WEEK1,CROP,QUAL,l in LOC,z in ZON:cidloc[l]==cidzone[z]} >=0;

var VSZD {WEEK1,CROP,QUAL,ZON,DC} >=0;

var VSZC {WEEK1,CROP,QUAL,ZON,CUST,MOD} >=0;

var VSDC {WEEK1,CROP,QUAL,DC,CUST,MOD} >= 0;

var VInvw {WEEK1,CROP,QUAL,ZON} >=0;

var VInvd {WEEK1,CROP,QUAL,DC} >=0;  # Inventory at the DC

param VInvw_Output{1..nCUT,WEEK1,CROP,QUAL,ZON};

param VInvd_Output{1..nCUT,WEEK1,CROP,QUAL,DC};

### Softening of constraints
```

```
var Add_War {ZON} >= 0;
var VAlpha >=0;
var theta;
param max_dual_value {CUTS,SCEN};
param max_dual {CUTS,SCEN};
param max_dual_theta_r {CUTS};
param max_dual_theta_T_1 {CUTS,WEEKP,CROP,LOC,TECH};
param max_dual_theta_T_2 {CUTS,WEEKP,CROP,LOC,TECH};
param temp;
param max_dual_temp;
param cut_coeff{CUTS,SCEN};
param alpha {CUTS,SCEN};
param beta {CUTS,SCEN,WEEKP,CROP,LOC,TECH};
param temp_mat{SCEN,CUTS};
####################
# MASTER PROBLEM #
# Objective function
maximize first_stage:
- sum{p in WEEKP,j in CROP,l in LOC,u in TECH} (Cplant[j]*VPlant[p,j,l,u])
# Planting Costs (at this point assuming that it does not vary based on technology
- sum {p in WEEKP, z in ZON, u in TECH, j in CROP, l in LOC: cidloc[l]==cidzone[z]}
(Coper[u,z]+Ctech[u,z])*VPlant[p,j,l,u]   # Becomes the technology costs
```

```
- theta;
# Constraints
subject to TechInvL:          # Investment in Technologies is limited by available capital
sum{z in ZON, j in CROP, l in LOC,   u in TECH: cidloc[l]== cidzone[z]}
VB[j,l,u]*Ctech[u,z] <= AvailCap;
subject to Tot_land {l in LOC, j in CROP, u in TECH}:
sum{p in WEEKP} VPlant[p,j,l,u] <= LAvail[l]*VB[j,l,u];
subject to Tot_land_Loc {l in LOC}:
sum{p in WEEKP,j in CROP, u in TECH} VPlant[p,j,l,u] <= LAvail[l];
subject to TechType{j in CROP}:
sum{u in TECH,l in LOC} VB[j,l,u] <=5;
subject to M_Prod {p in WEEKP, j in CROP, l in LOC, u in TECH}:
VPlant[p,j,l,u] <= maxl[j]*  VB[j,l,u];
subject to Min_Prod {j in CROP, l in LOC}:
sum{p in WEEKP,u in TECH} VPlant[p,j,l,u] >= sum{u in TECH} minl[j]*VB[j,l,u];
##### Optimality cuts
subject to Cut_Defn {s in SCEN}:
theta >=  if s == cur_scen then
cut_coeff[nCUT,cur_scen] * #cut_coeff[nCUT,cur_scen]*
(  sum{k in SCEN}
 (
```

sum{(p,h,z) in WEEK5} Water_Allocation_dual[max_dual[nCUT,k],p,h,z]*-1*LRainRec[k,p,h,z]

+ sum{h in WEEKH, j in CROP, q in QUAL,l in LOC, z in ZON: cidloc[l]==cidzone[z] and q==2}Shipment_L_dual[max_dual[nCUT,k],h,j,q,l,z]*COL[h,j,q]/25

+ sum{h in WEEKH,j in CROP,q in QUAL,z in ZON} Initial_hold_W_dual[max_dual[nCUT,k],h,j,q,z]

+ sum{(h,t) in WEEK1,j in CROP,q in QUAL,z in ZON: t>h} Invent_W_dual[max_dual[nCUT,k],h,t,j,q,z]

+ sum{h in WEEKH,j in CROP,q in QUAL,d in DC} Initial_hold_DC_dual[max_dual[nCUT,k],h,j,q,d]

+ sum{(h,t) in WEEK1,j in CROP,q in QUAL,d in DC: t>h} Invent_DC_dual[max_dual[nCUT,k],h,t,j,q,d]

+ sum{h in WEEKH,j in CROP,q in QUAL,z in ZON:q==2} Tot_packaging_dual[max_dual[nCUT,k],h,j,q,z]*COL[h,j,q]

+ sum{j in CROP,i in CUST,t in WEEKS} Max_Demand_dual[max_dual[nCUT,k],j,i,t]*MaxDem[t,j,i]

+ sum{j in CROP,i in CUST,t in WEEKS} Min_Demand_dual[max_dual[nCUT,k],j,i,t] *-1*MinDem[t,j,i]

+ sum{z in ZON, h in WEEKH} Ware_Z_Cap_dual[max_dual[nCUT,k],z,h]*WZ_Cap[z]

+ sum{j in CROP,q in QUAL,z in ZON, d in DC} Ship_const_Z_dual[max_dual[nCUT,k],j,q,z,d]

+ sum {j in CROP,q in QUAL,z in ZON, i in CUST, m in MOD}

```
Ship_const_W_dual[max_dual[nCUT,k],j,q,z,i,m]
+ sum {j in CROP,q in QUAL,d in DC, i in CUST, m in MOD}
Ship_const_DC_dual[max_dual[nCUT,k],j,q,d,i,m]
- sum{(p,h) in WEEK6, j in CROP, z in ZON, u in TECH, l in LOC:cidloc[l]==cidzone[z]}
(Micro_harvest_dual[max_dual[nCUT,k],h,j,z,u,l]*YDist[p,h,j,z,u]
*yield_sub[k,p,j,z,u]*VPlant[p,j,l,u])
- sum{(p,h,z) in WEEK5,l in LOC, u in TECH, j in CROP: cidloc[l]==cidzone[z]}
(Water_Allocation_dual[max_dual[nCUT,k],p,h,z]* MinWReq[j,u] * VPlant[p,j,l,u]))
- sum {p in WEEKP,j in CROP,l in LOC,u in TECH} (Cplant[j]*VPlant[p,j,l,u])
# Planting Costs (at this point assuming that it does not vary based on technology
- sum {p in WEEKP, z in ZON, u in TECH, j in CROP, l in LOC: cidloc[l]==cidzone[z]}
(Coper[u,z]+Ctech[u,z])*VPlant[p,j,l,u])
 else
alpha[nCUT,s] - sum{p in WEEKP, j in CROP, l in LOC, u in TECH}
beta[nCUT,s,p,j,l,u]*VPlant[p,j,l,u]*1;
# Planting Costs (at this point assuming that it does not vary based on technology
#################
# SUBPROBLEM
# Objective function
maximize second_stage:(
sum {(h,t) in WEEK1, j in CROP, q in QUAL,i in CUST,d in DC, m in MOD}
VSDC[h,t,j,q,d,i,m]*price[t,j,i]   ##Selling to customer directly from packing facility
```

```
-sum{z in ZON,h in WEEKH,j in CROP, q in QUAL} (VPACK[h,j,q,z]/25*Ccase[j])
# Packing costs
-sum{(h,t) in WEEK1, j in CROP, q in QUAL, z in ZON} VInvw[h,t,j,q,z]*Cw[z]
-sum{(h,t) in WEEK1, j in CROP, q in QUAL, d in DC} VInvd[h,t,j,q,d]*Cd[d]
-sum{ (p,h,z) in WEEK5} (VWatAll[p,h,z]*Cwater[z])
# Transportation costs
-sum {(h,t) in WEEK1, j in CROP,q in QUAL,d in DC, i in CUST,m in MOD}
VSDC[h,t,j,q,d,i,m]*CTDC[d,i,m]          # From facility to customer
-sum {(h,t) in WEEK1, j in CROP,q in QUAL,z in ZON,d in DC}
VSZD[h,t,j,q,z,d]*CTZD[z,d]          # From facility to customer
-sum {(h,t) in WEEK1, j in CROP,q in QUAL,l in LOC, z in ZON:cidloc[l]==cidzone[z]}
VSLZ[h,t,j,q,l,z]*CTLZ[z]
#### Keep feasibility
-sum{z in ZON} 10000000*Add_War[z]);
#######
# Constraints
##Resource Allocation
subject to Water_Allocation {(p,h,z) in WEEK5}:
VWatAll[p,h,z] >= -LRainRec[cur_scen,p,h,z] + sum{l in LOC, u in TECH, j in CROP:
cidloc[l]==cidzone[z]}  ( MinWReq[j,u] * plant_master[p,j,l,u] );
# Obtains the harvested product at a micro level
```

```
subject to Micro_harvest {h in WEEKH, j in CROP, z in ZON, u in TECH, l in LOC:cidloc[l]==cidzone[z]}:    # Limit harvest by amount planted #
VMicroHarv[h,j,l,u] =    sum{p in WEEKP:(p,h) in WEEK6} plant_master[p,j,l,u] *YDist[p,h,j,z,u] * yield_sub[cur_scen,p,j,z,u];
#Logistic restrictions
subject to Shipment_L {h in WEEKH, j in CROP, q in QUAL,l in LOC, z in ZON: cidloc[l]==cidzone[z]}:
VSLZ[h,h,j,q,l,z] = sum{u in TECH:q==2 and cidloc[l]==cidzone[z]}
COL[h,j,q]*VMicroHarv[h,j,l,u];
#Logistic restrictions
subject to Tot_packaging {h in WEEKH,j in CROP,q in QUAL,z in ZON}:
VPACK[h,j,q,z] = sum{l in LOC,u in TECH:q==2 and cidloc[l]==cidzone[z]}
COL[h,j,q]*VMicroHarv[h,j,l,u]/25;
subject to Ware_Z_Cap {z in ZON, h in WEEKH}:
sum{j in CROP, q in QUAL: q==2} VInvw[h,h,j,q,z]/25 <=WZ_Cap[z]+Add_War[z];
#Initial shipment to warehouse at each of the zones
 subject to Initial_hold_W {h in WEEKH,j in CROP,q in QUAL,z in ZON}:
VInvw[h,h,j,q,z] = sum{l in LOC: cidloc[l]==cidzone[z]} VSLZ[h,h,j,q,l,z]- sum{d in DC} VSZD[h,h+TimeZD[z,d],j,q,z,d];
 # Inventory at the warehouses
subject to Invent_W {(h,t) in WEEK1,j in CROP,q in QUAL,z in ZON: t>h}:
VInvw[h,t,j,q,z] = VInvw[h,t-1,j,q,z] - sum{d in DC} VSZD[h,t+TimeZD[z,d],j,q,z,d];
```

```
#Initial shipment to warehouse at each of the zones
subject to Initial_hold_DC {h in WEEKH,j in CROP,q in QUAL,d in DC}:
VInvd[h,h,j,q,d] = sum{z in ZON} VSZD[h,h,j,q,z,d]- sum{i in CUST,m in MOD}
VSDC[h,h,j,q,d,i,m];
 # Inventory at the warehouses
subject to Invent_DC {(h,t) in WEEK1,j in CROP,q in QUAL,d in DC:t>h}:
VInvd[h,t,j,q,d] = VInvd[h,t-1,j,q,d] +sum{z in ZON} VSZD[h,t,j,q,z,d] -sum{i in
CUST,m in MOD} VSDC[h,t,j,q,d,i,m];
# Capacity at the warehouse
subject to Ship_const_Z {j in CROP, q in QUAL, z in ZON, d in DC}:
sum{h in WEEKH,t in WEEKS: t > h + TimeZD[z,d] and (h,t) in WEEK1}
VSZD[h,t,j,q,z,d] + sum{h in WEEKH,t in WEEKS: t < h + TimeZD[z,d] and (h,t) in
WEEK1} VSZD[h,t,j,q,z,d] <=0;
subject to Ship_const_W {j in CROP, q in QUAL, z in ZON, i in CUST, m in MOD}:
sum{h in WEEKH,t in WEEKS: t > h + TimeZC[z,i,m] and (h,t) in WEEK1}
VSZC[h,t,j,q,z,i,m] + sum{h in WEEKH,t in WEEKS: t < h + TimeZC[z,i,m] and (h,t) in
WEEK1} VSZC[h,t,j,q,z,i,m] <=0;
subject to Ship_const_DC {j in CROP, q in QUAL, d in DC, i in CUST, m in MOD}:
sum{h in WEEKH,t in WEEKS: t > h + TimeDC[d,i,m] and (h,t) in WEEK1}
VSDC[h,t,j,q,d,i,m] + sum{h in WEEKH,t in WEEKS: t < h + TimeDC[d,i,m] and (h,t) in
WEEK1} VSDC[h,t,j,q,d,i,m] <=0;
subject to Inv_Const_W:
```

```
sum{j in CROP, q in QUAL, z in ZON,(h,t) in WEEK1: t > 53} VInvw[h,t,j,q,z]  <=0;
subject to Inv_Const_DC:
sum{j in CROP, q in QUAL, d in DC, (h,t) in WEEK1: t > 53} VInvd[h,t,j,q,d]  <=0;
#Demand
subject to Max_Demand {j in CROP,i in CUST,t in WEEKS}:          Modified. Took out
Warehouse, only considered DC
sum{h in WEEKH, d in DC, q in QUAL,m in MOD: h + SL[j] >= t >= h and q<=Qmin[i]
and (h,t) in WEEK1} VSDC[h,t,j,q,d,i,m]   <= MaxDem[t,j,i];
subject to Min_Demand {j in CROP,i in CUST,t in WEEKS}:
sum{h in WEEKH, d in DC, q in QUAL,m in MOD: h + SL[j] >= t >= h and q<=Qmin[i]
and (h,t) in WEEK1} VSDC[h,t,j,q,d,i,m]   >= MinDem[t,j,i];
```

```
#####
## R-SCRIPT
####
start_tables_SD <- function(){
Water_Allocation_dual_table  <- matrix(0,ncol = 5);
Water_Allocation_dual_table    [1,]    <-    c('CUTS',    'WEEKP',    'WEEKH',    'ZON',
'Water_Allocation_dual');
 write.table(Water_Allocation_dual_table      ,file      =      '../../../AMPLcml/ID/DUALS/
water_allocation_dual.csv',sep=',', row.names=FALSE,col.names = FALSE)
micro_harvest_dual_table <- matrix(0,ncol = 7);
```

```
micro_harvest_dual_table[1,] <- c('CUTS','WEEKH','CROP',
'ZON','TECH','LOC','Micro_harvest_dual');
write.table(micro_harvest_dual_table,file = '../../../AMPLcml/ID/DUALS/
micro_harvest_dual.csv',sep=',', row.names=FALSE,col.names = FALSE)
Tot_packaging_dual_table <- matrix(0,ncol = 6);Tot_packaging_dual_table[1,] <-
c('CUTS','WEEKH','CROP','QUAL','ZON','Tot_packaging_dual');
write.table(Tot_packaging_dual_table,file = '../../../AMPLcml/ID/DUALS/
tot_packaging_dual.csv',sep=',', row.names=FALSE,col.names = FALSE)
Shipment_L_dual_table <- matrix(0,ncol = 7);Shipment_L_dual_table[1,] <- c('CUTS',
'WEEKH', 'CROP', 'QUAL','LOC','ZON','Shipment_L_dual');
write.table(Shipment_L_dual_table,file = '../../../AMPLcml/ID/DUALS/
shipment_l_dual.csv',sep=',', row.names=FALSE,col.names = FALSE)
Max_Demand_dual_table <- matrix(0,ncol = 5);
Max_Demand_dual_table[1,] <- c('CUTS','CROPS',
'CUST','WEEKS','Max_Demand_dual');
write.table(Max_Demand_dual_table,file = '../../../AMPLcml/ID/DUALS/
max_dem_dual.csv',sep=',', row.names=FALSE,col.names = FALSE)
Min_Demand_dual_table <- matrix(0,ncol = 5);
Min_Demand_dual_table[1,] <- c('CUTS','CROPS',
'CUST','WEEKS','Min_Demand_dual');
write.table(Min_Demand_dual_table,file = '../../../AMPLcml/ID/DUALS/
min_dem_dual.csv',sep=',', row.names=FALSE,col.names = FALSE)
```

```
Ware_Z_Cap_dual_table <- matrix(0,ncol = 4);
Ware_Z_Cap_dual_table[1,] <- c('CUTS','ZON','WEEKH','Ware_Z_Cap_dual');
write.table(Ware_Z_Cap_dual_table,file = '../../../AMPLcml/ID/DUALS/
ware_z_cap_dual.csv',sep=',', row.names=FALSE,col.names = FALSE)
Invent_W_dual_table <- matrix(0,ncol = 7);Invent_W_dual_table[1,] <-
c('CUTS','WEEKH','WEEKS', 'CROP','QUAL','ZON','Invent_W_dual');
write.table(Invent_W_dual_table,file = '../../../AMPLcml/ID/DUALS/
invent_w_dual.csv',sep=',', row.names=FALSE,col.names = FALSE)
Invent_DC_dual_table <- matrix(0,ncol = 7);Invent_DC_dual_table[1,] <-
c('CUTS','WEEKH','WEEKS','CROP', 'QUAL','DC','Invent_DC_dual');
write.table(Invent_DC_dual_table,file = '../../../AMPLcml/ID/DUALS/
invent_dc_dual.csv',sep=',', row.names=FALSE,col.names = FALSE)
Initial_hold_W_dual_table <- matrix(0,ncol = 6);Initial_hold_W_dual_table[1,] <-
c('CUTS','WEEKH','CROP','QUAL','ZON','Initial_hold_W_dual');
write.table(Initial_hold_W_dual_table,file = '../../../AMPLcml/ID/DUALS/
initial_hold_w_dual.csv',sep=',', row.names=FALSE,col.names = FALSE)
Initial_hold_DC_dual_table <- matrix(0,ncol = 6);Initial_hold_DC_dual_table[1,] <-
c('CUTS','WEEKH','CROP','QUAL','DC','Initial_hold_DC_dual');
write.table(Initial_hold_DC_dual_table,file = '../../../AMPLcml/ID/DUALS/
initial_hold_dc_dual.csv',sep=',', row.names=FALSE,col.names = FALSE)
Ship_const_Z_dual_table <- matrix(0,ncol = 6);Ship_const_Z_dual_table [1,] <-
c('CUTS','CROP','QUAL','ZON','DC','Ship_const_Z_dual');
```

```
write.table(Ship_const_Z_dual_table,file = '../../../AMPLcml/ID/DUALS/ship_const_z_dual.csv',sep=',', row.names=FALSE,col.names = FALSE)
Ship_const_W_dual_table <- matrix(0,ncol = 7);
Ship_const_W_dual_table[1,] <- c('CUTS','CROP','QUAL','ZON','CUST','MOD','Ship_const_W_dual');
write.table(Ship_const_W_dual_table,file = '../../../AMPLcml/ID/DUALS/ship_const_w_dual.csv',sep=',', row.names=FALSE,col.names = FALSE)
Ship_const_DC_dual_table <- matrix(0,ncol = 7);Ship_const_DC_dual_table[1,] <- c('CUTS','CROP','QUAL','DC','CUST','MOD','Ship_const_DC_dual');
write.table(Ship_const_DC_dual_table,file = '../../../AMPLcml/ID/DUALS/ship_const_dc_dual.csv',sep=',', row.names=FALSE,col.names = FALSE)
Max_Profits_dual_table <- matrix(0,ncol = 2);
Max_Profits_dual_table[1,] <- c('CUTS','Max_Profits_dual');
write.table(Max_Profits_dual_table,file = '../../../AMPLcml/ID/DUALS/max_profits_dual.csv',sep=',', row.names=FALSE,col.names = FALSE)
max_dual_theta_r_table <- matrix(0, ncol=2);
max_dual_theta_r_table[1,] <- c('CUTS','max_dual_theta_r');
write.table(max_dual_theta_r_table ,file= '../../../AMPLcml/ID/AMPL_RUN/max_dual_theta_r.csv',sep=',', row.names=FALSE,col.names = FALSE)
max_dual_theta_T_1_table <- matrix(0, ncol=6);
max_dual_theta_T_1_table[1,] <- c('CUTS','WEEKP','CROP','LOC','TECH','max_dual_theta_T_1');
```

```
write.table(max_dual_theta_T_1_table ,file= '../../../AMPLcml/ID/AMPL_RUN/
max_dual_theta_T_1.csv',sep=',', row.names=FALSE,col.names = FALSE)
max_dual_theta_T_2_table <- matrix(0, ncol=6);
max_dual_theta_T_2_table[1,] <-
c('CUTS','WEEKP','CROP','LOC','TECH','max_dual_theta_T_2');
write.table(max_dual_theta_T_2_table ,file= '../../../AMPLcml/ID/AMPL_RUN/
max_dual_theta_T_2.csv',sep=',', row.names=FALSE,col.names = FALSE)
write.csv(NULL,file='../../../AMPLcml/ID/AMPL_RUN/
Vplant_master_plot.csv',row.names = FALSE)
write.csv(NULL,file='../../../AMPLcml/ID/AMPL_RUN/theta_values_sd.csv',row.names
= FALSE)
t <- data.frame(1);
names(t) <- c('Master');write.csv(t,file='../../../AMPLcml/ID/AMPL_RUN/
first_stage_value.csv',row.names = FALSE);

write.csv(NULL,file='../../../AMPLcml/ID/AMPL_RUN/
plant_master_plot.csv',row.names = FALSE)
write.csv(NULL,file='../../../AMPLcml/ID/OUTPUT/vmicroharv_SD.csv',row.names =
FALSE)
write.csv(NULL,file='../../../AMPLcml/ID/OUTPUT/vmicroharv_SD_hist.csv',row.name
s = FALSE)
write.csv(NULL,file='../../../AMPLcml/ID/OUTPUT/vslz_SD.csv',row.names = FALSE)
```

```
write.csv(NULL,file='../../../AMPLcml/ID/OUTPUT/vszc_SD.csv',row.names = FALSE)
write.csv(NULL,file='../../../AMPLcml/ID/OUTPUT/vsdc_SD.csv',row.names = FALSE)
write.csv(NULL,file='../../../AMPLcml/ID/OUTPUT/vinvw_stoch_SD.csv',row.names =
FALSE)
write.csv(NULL,file='../../../AMPLcml/ID/OUTPUT/vinvd_stoch_SD.csv',row.names =
FALSE)
write.csv(NULL,file = '../../../AMPLcml/ID/AMPL_RUN/yield_sub.csv')
write.csv(NULL,file = '../../../AMPLcml/ID/AMPL_RUN/cut_dfn.txt')
### Sample SD
write.csv(NULL,file = '../../../AMPLcml/ID/AMPL_RUN/k.csv')
sec_const_dual_table <- matrix(0,ncol = 3);sec_const_dual_table [1,] <-
c('SCEN','Dim','sec_const_dual');
write.table(sec_const_dual_table ,file = '../../../AMPLcml/ID/DUALS/
sec_const_dual_table.csv',sep=',', row.names=FALSE,col.names = FALSE)
max_dual_table_v3  <- matrix(0,ncol = 2);
max_dual_table_v3[1,] <- c('SCEN','max_dual');
write.table(max_dual_table_v3 ,file = '../../../AMPLcml/ID/AMPL_RUN/
max_dual_table_v3.csv',sep=',', row.names=FALSE,col.names = FALSE)
max_dual_table  <- matrix(0,ncol = 3);
max_dual_table[1,] <- c('CUTS','SCEN','max_dual');
write.table(max_dual_table ,file = '../../../AMPLcml/ID/AMPL_RUN/
max_dual_table.csv',sep=',', row.names=FALSE,col.names = FALSE)
```

```
max_dual_value_table  <- matrix(0,ncol =3);
max_dual_value_table[1,] <- c('CUTS','SCEN','max_dual_value');
write.table(max_dual_value_table ,file = '../../../AMPLcml/ID/AMPL_RUN/
max_dual_value_table.csv',sep=',', row.names=FALSE,col.names = FALSE)
alpha_table  <- matrix(0,ncol = 3);alpha_table[1,] <- c('CUTS','SCEN','alpha');
write.table(alpha_table ,file = '../../../AMPLcml/ID/AMPL_RUN/alpha_table.csv',sep=',',
row.names=FALSE,col.names = FALSE)
beta_table  <- matrix(0,ncol = 7);
beta_table[1,] <- c('CUTS','SCEN','WEEKP','CROP','LOC','TECH','beta');
write.table(beta_table ,file = '../../../AMPLcml/ID/AMPL_RUN/beta_table.csv',sep=',',
row.names=FALSE,col.names = FALSE)
alpha_table_ex  <- matrix(0,ncol = 3);alpha_table_ex [1,] <- c('ITER','SCEN','alpha');
write.table(alpha_table_ex,file = '../../../AMPLcml/ID/AMPL_RUN/
alpha_table_ex.csv',sep=',', row.names=FALSE,col.names = FALSE)
beta_table_ex  <- matrix(0,ncol = 3);beta_table_ex [1,] <- c('ITER','SCEN','beta');
write.table(beta_table_ex ,file = '../../../AMPLcml/ID/AMPL_RUN/
beta_table_ex.csv',sep=',', row.names=FALSE,col.names = FALSE)
temp_mat_table  <- matrix(0,ncol = 3);
temp_mat_table[1,] <- c('SCEN','CUTS','temp_mat');
write.table(temp_mat_table ,file = '../../../AMPLcml/ID/AMPL_RUN/
temp_mat.csv',sep=',', row.names=FALSE,col.names = FALSE)
}
```

```
### R-Script for Decomposition algorithm

rm(list=ls())

source('Crear_Sets_Tables.R')

source('Run_AMPL_SetUP.R')

#Load_Sets()

run_libraries()

start_tables_SD()

#####

version <- 'SD_V9'

cur_path <- dirname(dirname(dirname(getwd()))); path_to_directory <- cur_path;

cur_path <- gsub('/','\\\\',cur_path)

cur_path <- paste(cur_path,'AMPLcml','ID','Stochastic',version,sep='\\');

path_to_model <- cur_path;

cur_path <- paste('ampl',cur_path)

#########################

update_paths();


###

nCut <- 0

prices  <-  data.frame(read.csv(file  =  '../../../AMPLcml/ID/Datasets/PRICES_table.csv'));

names(prices)[ncol(prices)] <- c('price')

prcp <- data.frame(read.csv(file = '../../../AMPLcml/ID/Datasets/PRCP_scen_table.csv'));
```

```
prcp <- prcp[,-ncol(prcp)]
ZON <- data.frame(read.csv(file = '../../../AMPLcml/ID/Datasets/ZON_table.csv'))[,1]
TECH <- data.frame(read.csv(file = '../../../AMPLcml/ID/Datasets/TECH.csv'))[,1]
CROP <- data.frame(read.csv(file = '../../../AMPLcml/ID/Datasets/CROP_table.csv'))[,1]
############
yields_sample <- read.csv(file='../../../AMPLcml/ID/Datasets/Yields_sampled.csv')
yields.pca_transf <- read.csv(file='../../../AMPLcml/ID/Datasets/
Yields_PCA_transformed.csv')
#####
yield_sub <- NULL
prcp_sub_table <- NULL
prob_scen <- NULL
random_sample_scen <- NULL
master_values <- NULL
yield_sub.pca.transf <- NULL
scen_dist_yield <-NULL

random_sample_scen <- data.frame(sample(unique(yields_sample$ITER
[order(yields_sample$ITER)])))
prices_sample <- prices[prices$SCEN%in%random_sample_scen[,1],]
plant_master_table  <- matrix(0,ncol = 5);
plant_master_table[1,] <- c('WEEKP','CROP','LOC','TECH','plant_master');
```

```
write.table(plant_master_table        ,file        =        '../../../AMPLcml/ID/AMPL_RUN/
plant_master.csv',sep=',', row.names=FALSE,col.names = FALSE)
cor_mat <- NULL
prob_scen <- NULL
cut_coeffs_table <- NULL
max_dual_hist<- NULL
for(j in 1:nrow(random_sample_scen))  {
   if(j>147)   {     break;  }
    nCut <- j
   print(paste('CUT',nCut,sep = ' '))
  s <- random_sample_scen[j,1]
   test_solved <-'not_solved';
   t <- NULL;
   force_write_csv(t,'../../../AMPLcml/ID/AMPL_RUN/solve_result.csv')
   print(paste('SCENARIO',s,':',j/nrow(random_sample_scen)*100,'%',sep=' '))
 if(nCut==1)
 {      plant_master <- data.frame(read.csv('../../../AMPLcml/ID/AMPL_RUN/
          plant_master_iter_0.csv',header = TRUE))
 force_write_csv(plant_master,'../../../AMPLcml/ID/AMPL_RUN/plant_master.csv')
 }
  prices_scen <- prices[prices$SCEN==1&
   prices$CROP%in%CROP, ]
```

```
  prcp_scen <- prcp[prcp$SCEN==s,]
  #### iter is the same as scenario. Each iteration is actually a scenario.
  yields_scen <- yields_sample[yields_sample$ITER==s&
                    yields_sample$ZON%in%ZON&
                    yields_sample$CROP%in%CROP&
                    yields_sample$TECH%in%TECH,-1];
 yields_scen.agg <- aggregate(Yield~SCEN+WEEKP+CROP+
 ZON+TECH,yields_scen,sum)
 params <- data.frame(t(c(s,nCut))); names(params) <- c('cur_scen','nCUT')
 temp <- data.frame(t(c(s,0.1))); names(temp) <- c('SCEN','Prob')
prob_scen <- rbind(prob_scen,temp)
prob_scen <- prob_scen[!duplicated(prob_scen),]
write.csv(prob_scen,file='../../../AMPLcml/ID/AMPL_RUN/
prob_scen.csv',row.names = FALSE)
force_write_csv(params,'../../../AMPLcml/ID/AMPL_RUN/params_scen.csv')
force_write_csv(prices_scen,'../../../AMPLcml/ID/AMPL_RUN/prices_scen.csv')
force_write_csv(yields_scen,'../../../AMPLcml/ID/AMPL_RUN/yields_scen.csv')
## for master problem
names(yields_scen.agg)[ncol(yields_scen.agg)] <- c('yield_sub')
yield_sub <- rbind(yield_sub,yields_scen.agg[,
c('SCEN','WEEKP','CROP','ZON','TECH','yield_sub')])
yield_sub_table <- yield_sub[!duplicated(yield_sub),]
```

```
write.csv(yield_sub_table,file = '../../../AMPLcml/ID/AMPL_RUN/m
yield_sub.csv',row.names = FALSE)
prcp_sub_table <- rbind(prcp_sub_table,prcp_scen)
prcp_sub_table <- prcp_sub_table[!(duplicated(prcp_sub_table)),]
write.csv(prcp_sub_table,file = '../../../AMPLcml/ID/AMPL_RUN/
prcp_sub_table.csv',row.names = FALSE)
if(j>1)
{
plant_master_reordered <- data.frame(read.csv( '../../../AMPLcml/ID/AMPL_RUN/
plant_master.csv'))
plant_master_reordered <- plant_master_reordered[order(-plant_master_reordered$
plant_master),]
force_write_csv(plant_master_reordered,'../../../AMPLcml/ID/AMPL_RUN/plant_master
.csv')
}
write.csv(NULL,file='../../../AMPLcml/ID/OUTPUT/vszc_SD.csv',row.names = FALSE)
write.csv(NULL,file='../../../AMPLcml/ID/OUTPUT/vszd_SD.csv',row.names = FALSE)
write.csv(NULL,file='../../../AMPLcml/ID/OUTPUT/vsdc_SD.csv',row.names = FALSE)
write.csv(NULL,file='../../../AMPLcml/ID/OUTPUT/vmicroharv_SD.csv',row.names =
FALSE)
write.csv(NULL,file='../../../AMPLcml/ID/OUTPUT/vinvw_stoch_SD.csv',row.names =
FALSE)
```

```
write.csv(NULL,file='../../../AMPLcml/ID/OUTPUT/vinvd_stoch_SD.csv',row.names    =
FALSE)
write.csv(NULL,file='../../../AMPLcml/ID/AMPL_RUN/cut_dfn.txt')
test_solved <-'not_solved';
t <- NULL; force_write_csv(t,'../../../AMPLcml/ID/AMPL_RUN/solve_result.csv')
while(test_solved != 'solved')  ## solves ampl until it is able to correctly run
 {
     path_submodel <- paste(cur_path,'\\sub_model.mod',sep='')
     print(shell(path_submodel))
     test_solved <- read.csv('../../../AMPLcml/ID/AMPL_RUN/
    solve_result.csv',header = FALSE,stringsAsFactors = FALSE)
     test_solved <- substr(test_solved,nchar(test_solved)-5,nchar(test_solved))
     print(test_solved)
     t<- NULL;
     force_write_csv(t,'../../../AMPLcml/ID/AMPL_RUN/solve_result.csv')
  }
   order_duals()
   cut_coeffs_iter <- update_cut_coeffs(nCut,data.frame(prob_scen[,1]))
   cut_coeffs_table <- rbind(cut_coeffs_table,cut_coeffs_iter)
   cut_coeffs_table <- cut_coeffs_table[!duplicated(cut_coeffs_table),]

   force_write_csv(cut_coeffs_table,'../../../AMPLcml/ID/AMPL_RUN/cut_coeff.csv')
```

```
plant_master_table <- matrix(0,ncol = 5);plant_master_table[1,] <-
c('WEEKP','CROP','LOC','TECH','plant_master');
write.table(plant_master_table ,file = '../../../AMPLcml/ID/AMPL_RUN/
plant_master.csv',sep=',', row.names=FALSE,col.names = FALSE)
print('MASTER')
#### solve master problem
master_test_solved<-'not_solved';
write.csv(t,file='../../../AMPLcml/ID/AMPL_RUN/master_solve_result.csv');# clears the
solve display from ampl
while(master_test_solved != 'solved')  ## solves ampl until it is able to correctly run
{
    path_mastermodel <- paste(cur_path,'\\master_model.mod',sep='')
    print(shell(path_mastermodel))
    master_test_solved <- read.csv('../../../AMPLcml/ID/AMPL_RUN/
master_solve_result.csv',header = FALSE,stringsAsFactors = FALSE)
    master_test_solved <- substr(master_test_solved,nchar(master_test_solved)-5,
nchar(master_test_solved))
    print(master_test_solved)
    t<- NULL; force_write_csv(t,'../../../AMPLcml/ID/
AMPL_RUN/master_solve_result.csv')
 }
yields.pca_transf_sub <- yields.pca_transf[yields.pca_transf$ITER%in%prob_scen[,1],]
```

```
yields.pca_transf_not <- yields.pca_transf[!(yields.pca_transf$ITER%in%prob_scen[,1]),]
yields.pca_transf_sub$SELECT <- c('SELECTED')
yields.pca_transf_not$SELECT <- c('NOT_SELECTED')
# Graphs the master value's progression
   if(j>=4)
   {
     cut_plot <- plot_cut_max_dual(j)+
             theme_bw()

     yields.pca_transf_temp <- yields.pca_transf[yields.pca_transf$ITER==j,]
     yield_sub.pca.transf <- rbind(yield_sub.pca.transf,yields.pca_transf_temp)
values <- read.csv(file='../../../AMPLcml/ID/AMPL_RUN/ first_stage_value.csv',header = TRUE,stringsAsFactors = FALSE)
     values$iter <- seq.int(1,nrow(values))
     values$Master <- round(values$Master,digits = 0)
      ####### Application of SVM
     correlation <- data.frame(j);
     names(correlation)[1] <- c('iter')
     correlation$corr <- fit_svm_model(yields.pca_transf,j)

}
```

APPENDIX E

TECHONOLOGY ALLOCATION FORMULATION

```
# Data sets #
set WEEKP ordered;                         # weeks of the planning period
set WEEKH ordered;# weeks of the harvesting period
set CROP;                              # Crops for planting
set CUST;                               # Customers
set CON;
set FARMER;
set TECH;
set ZON;
set MOD;
set DC;
#####
set WEEK ordered;
set WEEKS ordered;
#Additional necessary sets to facilitate indexing
set WEEK2 within {WEEKP, WEEKH,CROP,TECH,ZON};
set WEEK6 within {WEEKP, WEEKH};
set WEEK1 within {WEEKH, WEEKS};
set WORK5 within {WEEKP, WEEK,CROP};
set WEEK5 within {WEEKP, WEEKH,ZON};
set QUAL:= 2..2;# Color characteristic of products
########
```

```
# Parameter definition #

# Production/Yield-Related

param YDist {WEEKP,WEEKH,CROP,TECH,ZON} >= 0;     # Expected YDist per crop

param Yield {WEEKP,CROP,TECH,ZON}>=0;   # Yield production expected

param LRainRec{WEEKP,WEEKH,ZON} >= 0;

param Cplant{CROP} >=0;

param Cwater{ZON}  >=0;                           # Cost of water per location l

param Cw {ZON} >= 0; # Cost of warehouse

param Cd {DC} >= 0; # Cost of DC

param Ccase{CROP} >=0;                            # Cost for packing case for crop j

param TraF{CROP}  >=0;

param WZ_Cap{ZON} >= 0;

param SL {CROP} >=0;                        # Shelf life of product k

param LT {CUST} >=0;                        # Lead time required by the customer

param Weight {CROP} >=0;  # Quantity in required of crop j to form a case of product k

param PZcap {ZON} >=0;             # Capacity of te packaging facility for a time period

param COL {WEEKH,CROP,QUAL}>=0;

param minl{CROP} >=0;                             # Minimum amount to plant per crop

param maxl{CROP} >=0;                             # Maximum amount to plant per crop

param CTZC{ZON,CUST,MOD} >=0;

param CTZD{ZON,DC} >=0;# Cost of transportation from zone z to customer c

param CTDC{DC,CUST,MOD} >=0;
```

```
param CTLZ{ZON} >=0;


param TimeZC {ZON,CUST,MOD} >=0;

param TimeDC {DC,CUST,MOD} >=0;

param TimeZD {ZON,DC} >=0;

param Ctech{TECH, ZON} >=0;                    # Cost of technology u for location l

param Coper{TECH, ZON} >= 0;

param MinWReq {CROP,TECH} >=0;

param Min_Prod {WEEKH,CROP,ZON} >= 0;

param MinUtil{FARMER} >= 0;

param LAvail {ZON,TECH} >=0;

param Clabor >= 0;

param Chire >= 0 ;

param Ctemp >= 0;

param MTemp >= 0;

param MFix >=0;

param LaborP {WEEKP,WEEK,CROP} >=0;

param LaborH {WEEKP,WEEK,CROP} >=0;

param cidfarmer {FARMER} symbolic;

param cidzone {ZON} symbolic;

# Market Related

param price {WEEKS,CROP,CUST} >= 0;
```

```
param CONPlant{CON,CROP,TECH} >=0;
param CONPrice {CON,CROP,TECH} >= 0;
param tot_farming_profit{f in FARMER};
param adjusted_profit;
param revenue_zone{j in CROP};
param investment_cost {j in CROP};
param logistic_cost {j in CROP};
########## Variable Definition
#Planting and resource decisions
var D_Plant {WEEKP,CROP,TECH,f in FARMER,CON,z in ZON:cidfarmer[f]==cidzone[z]} >=0;
var D_MicroHarv {WEEKH,CROP,TECH,f in FARMER,CON,z in ZON: cidfarmer[f]==cidzone[z]} >=0;
var D_CONSel {CON,FARMER,CROP,TECH} >=0 binary;
var OPL{WEEK,FARMER} >= 0;
var OPT{WEEK,FARMER} >= 0;
var HIRE{WEEK,FARMER} >= 0;
var FIRE{WEEK,FARMER} >= 0;
var SLZ {WEEK1,CROP,QUAL,f in FARMER,z in ZON: cidfarmer[f]==cidzone[z]} >=0;
var SZD {WEEK1,CROP,QUAL,ZON,DC} >=0;
var SDC {WEEK1,CROP,QUAL,DC,CUST,MOD} >= 0;
```

```
var Invw {WEEK1,CROP,QUAL,ZON} >=0;

var Invd {WEEK1,CROP,QUAL,DC} >=0;  # Inventory at the DC

var WatAll {WEEKP,WEEKH,ZON,CROP} >=0;

var Add_War{ZON} >=0;

######## Earnings

var D_farmer_earning {FARMER,CROP,TECH};

####################

# Objective Function #

maximize Yield_revenue:

# Market revenues

sum {(h,t) in WEEK1, j in CROP, q in QUAL,c in CUST,d in DC, m in MOD}
SDC[h,t,j,q,d,c,m]*price[t,j,c]#*1.5 ##Selling to customer directly from packing facility

- sum{h in WEEKH, u in TECH, r in CON, c in CUST, f in FARMER, j in CROP, z in
ZON:cidfarmer[f]==cidzone[z]}  price[h,j,c] * CONPrice[r,j,u] *D_MicroHarv[h,j,u,f,r,z]

-sum {p in WEEKP, r in CON, z in ZON, u in TECH, j in CROP, f in FARMER:
cidfarmer[f]==cidzone[z] and u=='IRR+CON'} (Ctech[u,z]*0.1)*D_Plant[p,j,u,f,r,z

-sum {p in WEEKP, r in CON, z in ZON, u in TECH, j in CROP, f in FARMER:
cidfarmer[f]==cidzone[z] and u=='IRR+GRH'} (Ctech[u,z]*0.1)*D_Plant[p,j,u,f,r,z]

-sum{(h,t) in WEEK1, j in CROP, q in QUAL, z in ZON} Invw[h,t,j,q,z]*Cw[z]

-sum{(h,t) in WEEK1, j in CROP, q in QUAL, d in DC} Invd[h,t,j,q,d]*Cd[d]

-sum {(h,t) in WEEK1, j in CROP,q in QUAL,d in DC, i in CUST,m in MOD}
SDC[h,t,j,q,d,i,m]*CTDC[d,i,m]          # From facility to customer
```

```
-sum {(h,t) in WEEK1, j in CROP,q in QUAL,z in ZON,d in DC}
SZD[h,t,j,q,z,d]*CTZD[z,d]          # From facility to customer



-sum{f in FARMER, t in WEEK} OPL[t,f]*Clabor
-sum{f in FARMER, t in WEEK} HIRE[t,f]*Chire
-sum{f in FARMER, t in WEEK} OPT[t,f]*Ctemp
;
#######
######################
##CONSTRAINTS
#
######
# Planting decisions
subject to Contract_Selection {u in TECH, z in ZON}:
sum{r in CON, p in WEEKP, f in FARMER, j in CROP:cidfarmer[f]==cidzone[z]}
D_Plant[p,j,u,f,r,z] <=  LAvail[z,u];
subject to Contract_Selection_Options {f in FARMER}:
sum{r in CON, u in TECH,j in CROP} D_CONSel[r,f,j,u] <= 1;
subject to Planting_Restriction {f in FARMER,z in ZON, r in CON,j in CROP,u in TECH:
cidfarmer[f]==cidzone[z]}:
sum{p in WEEKP} D_Plant[p,j,u,f,r,z] >= CONPlant[r,j,u]*D_CONSel[r,f,j,u];
```

```
subject to Water_Allocation {(p,h) in WEEK6, z in ZON,j in CROP}:
sum{f in FARMER, r in CON, u in TECH: cidfarmer[f]==cidzone[z]} ( MinWReq[j,u]
*D_Plant[p,j,u,f,r,z] ) <= LRainRec[p,h,z]  +  WatAll[p,h,z,j];


subject to Harv_Prod {h in WEEKH, f in FARMER,r in CON, u in TECH, j in CROP, z in
ZON: cidfarmer[f]==cidzone[z]}:
D_MicroHarv[h,j,u,f,r,z] <= sum{p in WEEKP:(p,h) in WEEK6} D_Plant[p,j,u,f,r,z] *
YDist[p,h,j,u,z] * Yield[p,j,u,z];# + test_harv_prod;
subject to Min_Cont_Inv{h in WEEKH, j in CROP, z in ZON}:
sum{f in FARMER, u in TECH, r in CON: cidfarmer[f]==cidzone[z]}
D_MicroHarv[h,j,u,f,r,z] >= Min_Prod[h,j,z];
subject to Util_Min{f in FARMER, z  in ZON: cidfarmer[f]==cidzone[z]}:
        sum{j in CROP, u in TECH} D_farmer_earning[f,j,u]
        -sum{t in WEEK} OPL[t,f]*Clabor
        -sum{t in WEEK} HIRE[t,f]*Chire
        -sum{t in WEEK} OPT[t,f]*Ctemp
        -sum {h in WEEKH, j in CROP, q in QUAL} SLZ[h,h,j,q,f,z]*CTLZ[z]
        -sum{(p,h,z) in WEEK5, j in CROP} (WatAll[p,h,z,j]*Cwater[z])
         >= sum{j in CROP, r in CON, u in TECH}
MinUtil[f]*D_CONSel[r,f,j,u]*CONPlant[r,j,u];
subject to Lab_Fields {h in WEEKH, f in FARMER, z in ZON: cidfarmer[f]==cidzone[z]}:
```

```
OPL[h,f]+OPT[h,f] >= sum{p in WEEKP, j in CROP, u in TECH, r in CON: (p,h,j) in
WORK5} (D_Plant[p,j,u,f,r,z]*(LaborP[p,h,j]+LaborH[p,h,j]))                    ;
subject to Shipment_L {h in WEEKH , j in CROP, q in QUAL, f in FARMER, z in ZON:
cidfarmer[f]==cidzone[z]}:
SLZ[h,h,j,q,f,z] = sum{u in TECH, r in CON:q==2} COL[h,j,q]*D_MicroHarv[h,j,u,f,r,z];
# Warehousing capacity at each ZON and DC
subject to Ware_Z_Cap {z in ZON, h in WEEKH}:
sum{j in CROP, q in QUAL: q==2} Invw[h,h,j,q,z]/25 <=WZ_Cap[z]+Add_War[z];
#Initial shipment to warehouse at each of the zones
subject to Initial_hold_W {h in WEEKH,j in CROP,q in QUAL,z in ZON}:
Invw[h,h,j,q,z] = sum{f in FARMER: cidfarmer[f]==cidzone[z]} SLZ[h,h,j,q,f,z]- sum{d
in DC} SZD[h,h+TimeZD[z,d],j,q,z,d];  # -sum{d in DC} SZD[h,h+TimeZD[z,d],j,q,z,d];
# Inventory at the warehouses
subject to Invent_W {(h,t) in WEEK1,j in CROP,q in QUAL,z in ZON: t>h}:
Invw[h,t,j,q,z] = Invw[h,t-1,j,q,z] - sum{d in DC} SZD[h,t+TimeZD[z,d],j,q,z,d];
#Initial shipment to warehouse at each of the zones
subject to Initial_hold_DC {h in WEEKH,j in CROP,q in QUAL,d in DC}:
Invd[h,h,j,q,d] = sum{z in ZON} SZD[h,h,j,q,z,d]- sum{i in CUST,m in MOD}
SDC[h,h,j,q,d,i,m];   #  - sum{i in CUST,m in MOD} SDC[h,h+TimeDC[d,i,m],j,q,d,i,m];
# Inventory at the warehouses
subject to Invent_DC {(h,t) in WEEK1,j in CROP,q in QUAL,d in DC:t>h}:
```

```
Invd[h,t,j,q,d] = Invd[h,t-1,j,q,d] +sum{z in ZON} SZD[h,t,j,q,z,d] -sum{i in CUST,m in MOD} SDC[h,t,j,q,d,i,m];
subject to Ship_const_Z {j in CROP, q in QUAL, z in ZON, d in DC}:
sum{(h,t) in WEEK1: t > h + TimeZD[z,d]} SZD[h,t,j,q,z,d] + sum{(h,t) in WEEK1: t < h + TimeZD[z,d]} SZD[h,t,j,q,z,d] =0;
subject to Ship_const_W {j in CROP, q in QUAL, z in ZON, i in CUST, m in MOD}:
sum{(h,t) in WEEK1: t > h + TimeZC[z,i,m]} SZC[h,t,j,q,z,i,m] + sum{(h,t) in WEEK1:t < h + TimeZC[z,i,m]} SZC[h,t,j,q,z,i,m] =0;

subject to Ship_const_DC {j in CROP, q in QUAL, d in DC, i in CUST, m in MOD}:
sum{(h,t) in WEEK1: t > h + TimeDC[d,i,m]} SDC[h,t,j,q,d,i,m] + sum{(h,t) in WEEK1: t < h + TimeDC[d,i,m]} SDC[h,t,j,q,d,i,m] =0;
subject to Ship_const_Z_Harv_Period:
sum{j in CROP, q in QUAL, d in DC,m in MOD, i in CUST, (h,t) in WEEK1: t >= 52} SDC[h,t,j,q,d,i,m]  + sum{j in CROP, q in QUAL, d in DC,m in MOD, i in CUST, (h,t) in WEEK1: t <= 13} SDC[h,t,j,q,d,i,m]<=0;
subject to Inv_Const_W:
sum{j in CROP, q in QUAL, z in ZON,(h,t) in WEEK1: t >= 52} Invw[h,t,j,q,z]  <=0;
subject to Inv_Const_DC:
sum{j in CROP, q in QUAL, d in DC, (h,t) in WEEK1: t >= 52} Invd[h,t,j,q,d]  <=0;
subject to Hire_init {t in WEEK,f in FARMER:t =first(WEEK)}:
HIRE[t,f]=OPL[t,f];
```

```
subject to Hire_Labor {t in WEEK,f in FARMER:29>t>first(WEEK)}:

HIRE[t,f]-FIRE[t,f]=OPL[t,f]-OPL[t-1,f];

subject to Hire_LabF {f in FARMER,t in WEEK: t>=28}:

HIRE[28,f]+ OPL[27,f]>=  OPL[t,f];

subject to Temporal {t in WEEK}:

sum {f in FARMER} OPT[t,f]<= MTemp;

subject to Fixed_lab:

sum {t in WEEK,f in FARMER} HIRE[t,f] <= MFix;

subject to Fire_Labor {t in WEEK,f in FARMER:t>29}:

OPL[t,f]-OPL[t-1,f]+FIRE[t,f]>=0;

subject to Labor_D_CON {f in FARMER, z in ZON: cidfarmer[f]==cidzone[z]}:

sum{t in WEEK} (OPL[t,f]+OPT[t,f]+FIRE[t,f]+HIRE[t,f]) <= MFix;

#### Farming output for earnings

subject to farmer_earning_const{f in FARMER,z in ZON,j in CROP, u in TECH:
cidfarmer[f]==cidzone[z]}:

D_farmer_earning[f,j,u] = sum{h in WEEKH, r in CON, c in CUST} price[h,j,c] *
CONPrice[r,j,u] * D_MicroHarv[h,j,u,f,r,z]

-sum{p in WEEKP, r in CON} (D_Plant[p,j,u,f,r,z]*Cplant[j])

-sum {p in WEEKP, r in CON: u=='IRR+CON'} (Coper[u,z]+
Ctech[u,z]*0.9)*D_Plant[p,j,u,f,r,z]# Becomes the technology costs

-sum {p in WEEKP, r in CON: u=='IRR+GRH'} (Coper[u,z]+
Ctech[u,z]*0.9)*D_Plant[p,j,u,f,r,z]#
```

;