PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
ESCUELA DE INGENIERÍA

# MULTISTAGE STOCHASTIC PROGRAMMING AS FLEXIBILITY SOURCE IN HIGHLY UNCERTAIN ENVIRONMENTS: ITS VALUE IN AN AGRICULTURE APPLICATION

**ELBIO LEONEL AVANZINI**

Thesis submitted to the Office of Graduate Studies in partial
fulfillment of the requirements for the degree of
Doctor en Ciencias de la Ingeniería

Advisor:
JORGE VERA ANDREO
ALEJANDRO MAC CAWLEY VERGARA

Santiago de Chile, (March, 2022)

PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE

SCHOOL OF ENGINEERING

# MULTISTAGE STOCHASTIC PROGRAMMING AS FLEXIBILITY SOURCE IN HIGHLY UNCERTAIN ENVIRONMENTS: ITS VALUE IN AN AGRICULTURE APPLICATION

**ELBIO LEONEL AVANZINI**

Members of the Committee:

| | |
|---|---|
| **JORGE VERA** | DocuSigned by: Jorge Vera A. 1CB4B1FD5F834A2... |
| **ALEJANDRO MAC CAWLEY** | DocuSigned by: Alejandro Mac Cawley V. 20DDCAC971A0463... |
| **VICTOR ALBORNOZ** | DocuSigned by: Victor Albornoz S. 2AD20F93A47347D... |
| **SERGIO MATURANA** | DocuSigned by: Sergio Maturana B433B3509ACC44B... |
| **GUSTAVO ANGULO** | DocuSigned by: Gustavo Angulo O. A4BF30B0D9C3411... |
| **LLUIS M. PLA ARAGONES** | DocuSigned by: Luis Miguel Pla Aragones FDA1CBE44D13452... |
| **GUSTAVO LAGOS** | DocuSigned by: Gustavo Lagos C. B558E6FD3B4641F... |

Thesis submitted to the Office of Graduate Studies in partial fulfillment of the requirements for the Degree Doctor in Engineering Sciences

Santiago de Chile, March, 2022

Thesis submitted to the Office of Graduate Studies in partial fulfillment of the requirements for the degree of
Doctor en Ciencias de la Ingeniería

Santiago de Chile, (March, 2022)

*To my family, Verónica, Tomás, y Agustín*

## ACKNOWLEDGEMENTS

I want to thank those who have actively contributed to achieving the goal of this cycle.

To my supervisors, Jorge Vera and Alejandro Mac Cawley, who have given their time and patience to guide the process. I add my special thanks to Alejandro, who has also supported me with numerous gestures that I consider signs of friendship. In the same way to his family who have received me in a kind and loving way. To the Evaluation Committee, who have facilitated the improvement of this work by making their criticisms in a relaxed atmosphere, without losing technical rigor. To Sergio Maturana, who has shared his experience in a selfless way. To Karina Norambuena, who has facilitated my activity in the university from the beginning. To Fernanda Kattan and her team who have actively supported me to achieve experiences that have a permanent stamp on me.

I want to thank my friend Carlos Monardes, whom I met during the doctorate, and with whom we share a valuable and growing friendship. Also, Lluis and Esteve, and those coffees that we drink so many times. It was a technical and life experience. A very special trip, and with whom I share a friendship and the desire to return now that we are more.

I also want to thank my brother Diego, my sister-in-law Pamela, and my nephews, who supported us since before coming to Chile and have always been with us. To my father Alfonso, Miguel and Alcira, my in-laws; it was not easy for them, but they supported us even with what that meant. And they still do.

Finally to Veronica and my little ones. The decision to leave Argentina was complex; but in the difficulty Veronica has supported me, helped and accompanied me. We share the adventure of coming, and now, that of our children, Tomás and Agustín, who have given their time to help me reach the goal. This is for them, looking for new alternatives for a healthier family life; Today we hope to see that sowing, and it will be with them. Thank you very much Vero, Tomi and Agustín, this is by and for you.

# TABLE OF CONTENTS

## LIST OF FIGURES

## LIST OF TABLES

## ABSTRACT

In this work, we have faced a critical problem in winemaking, harvesting task, characterized by high uncertainty. This thesis focuses on the decision-making process from the point of view of operations research, helping to bridge the gap in the literature where stochasticity has not been directly addressed.

We introduce multi-stage stochastic programming for the harvest process by considering different grape qualities, with quality degradation curves that depend on time and climatic conditions, the variable event. Decisions are about labor and its allocation to different lots to advance the task. The labor force may have different flexibility to cope with the rain, partially maintaining its productivity during dry periods. Stochastic multi-stage models also present a form of flexibility corresponding to the time when decisions are made. Both flexibilities are evaluated through the impact they generate on economic performance. The models that support decisions, multi-stage stochastic programming, and the traditional expected value problem, are compared. All of this is evaluated for various uncertainty conditions.

The computational effort, measured as the optimization time, appears as an obstacle to implement the multi-period model. To address this problem, we proposed the rolling tree algorithm, which resulted in a competitive tool. We test different configurations in the time structure and arrive at some conclusions that shed light on the relative weights in the expected gap versus the stochastic multi-period option, both in time and economic terms. We also analyze a notion of risk for the rolling tree approach.

This thesis contributes to filling the literature gap on the stochastic approach in fresh products, especially in agriculture and wine grapes. We also determine the impact of different types of flexibility under conditions of uncertainty, finding patterns that help in the choice of models to support management, focusing on economic performance. These types of conclusions are welcome, given that due to the multidimensional nature of flexibility and its difficulty in being measurable, they are difficult to reach. In the same vein, the rolling tree approach proposal offers a new model that maintains the advantages of

stochastic multi-period but with less computational effort, giving greater versatility to stochastic modeling.

**Keywords**: multistage stochastic programming, flexibility measures, rolling tree approach, computational effort, trees comparative, agriculture, fresh product, wine grape.

## RESUMEN

En este trabajo, nos hemos enfrentado a un problema crítico en la elaboración del vino, la labor de vendimia caracterizada por su alta incertidumbre. Esta tesis se centra en el proceso de toma de decisiones desde el punto de vista de la investigación operativa, contribuyendo a acortar el vacío de la literatura, donde la estocasticidad no se ha abordado de forma directa.

Introducimos la programación estocástica multietapa para el proceso de vendimia mediante, considerando diferentes calidades de uva, con curvas de degradación que dependen del tiempo y las condiciones climáticas, que es el evento variable. Las decisiones son sobre mano de obra y su asignación a distintos lotes para avanzar en la tarea. La mano de obra puede tener distinta flexibilidad para enfrentar la lluvia, manteniendo parcialmente su productividad de periodos secos. Los modelos multietapa estocástico, también presenta una forma de flexibilidad, correspondiente a la época en la que se toman las decisiones. Ambas flexibilidades son evaluadas a través del impacto que generan sobre el sobre el desempeño económico. Los modelos que soportan las decisiones y son comparados son la programación estocástica de varias etapas y el problema del valor esperado tradicional. Todo esto es evaluado para diversas condiciones de incertidumbre.

El esfuerzo computacional, medido como el tiempo de optimización, aparece como un obstáculo para implementar el modelo multiperiodo,. Para enfrentar este problema, propusimos el algoritmo de árbol rodante, que resultó en una herramienta competitiva. Repasamos diferentes configuraciones en la estructura temporal y llegamos a algunas conclusiones que arrojan luz sobre los pesos relativos en la brecha esperada frente a la opción multiperiodo estocásticas, tanto en términos de tiempo como económicos. También analizamos una noción de riesgo para la aproximación de árbol rodante.

Esta tesis contribuye a llenar el vacío de la literatura sobre el enfoque estocástico en productos frescos, especialmente en agricultura y en uva de vinificación. También determinamos el impacto de diferentes tipos de flexibilidad en condiciones de incertidumbre, encontrando patrones que ayudan en la elección de los modelos para apoyo a la gestión, enfocándonos en el desempeño económico. Este tipo de conclusiones son bienvenidas, dado que, por el carácter multidimensional de la flexibilidad y su dificultad para ser mensurable, son difíciles de alcanzar. En la misma línea, la propuesta del enfoque rolling tree, ofrece un nuevo modelo que mantiene las ventajas de multiperiodo estocástico, pero con menor esfuerzo computacional, dando mayor versatilidad al modelado estocástico.

**Palabras Claves**: Programación estocástica multietapa, medidas de flexibilidad, árbol de horizonte rodante, tiempo computacional, comparativa de árboles de decisión, agricultura, producto fresco, uva de vinificación..

## ACRONYMS

AO, appellation of origin

EA, Expected Value Problem when decisions are made after uncertainty is revealed

EV,EB, Expected Value Problem when decisions are made before uncertainty is revealed

EEV, expected value of the solution using the EV solution

EIV, expected input value

ES, Expected Shortfall

ESSV, expected skeleton solution value

EVA, the expected value of the solution of the expected value problem when some decisions are made after the uncertainty reveals itself

EVB, the expected value of the solution of the expected value problem when some decisions are made after the uncertainty reveals itself

EVPI, Expected Value of Perfect Information

EVR, Expected value of the solution of RTA

EVRE, Expected value of the solution of RTA when simplification is EV

EVRW, Expected value of the solution of RTA when simplification is WC

HSM, ripening pattern "premium-standard-medium"

LUDS, Loss of upgrading the deterministic solution

LUSS, loss using the skeleton solution

MA, multistage stochastic model where some decisions are made after uncertainty is revealed

MB, multistage stochastic model where some decisions are made before uncertainty is revealed

MC, multistage stochastic model where some decisions are made before uncertainty is revealed with the option to modify them later

MI, maturity index

MILP, mixed integer linear programming,

MS, multi-stage stochastic optimization, multi-stage stochastic programming

MSH, ripening pattern “medium-standard-premium”

MSV, Marginal Stage Value

NAC, non-anticipativity constraints

RHA, rolling horizon approach

RHVRS, rolling-horizon value of the reference scenario,

RP, Stochastic recourse problem

RTA, rolling tree algorithm or approximation

RTE, rolling tree model with expected value simplification

RTW, rolling tree model with worst case simplification

SAA, Sample Average Approximation

SOD, ripening pattern “same optimum day”

SP, Stochastic programming

TS, Time structure

VoHR, Value of rolling horizon

VSS, Value of Stochastic Solution

WA, wait-and-see model where some decisions are made after uncertainty is revealed

WB, wait-and-see model where some decisions are made before uncertainty is revealed

WC, wait-and-see model where some decisions are made before uncertainty is revealed with the option to modify them later

WD, Wasserstein Distance

WS, Wait-and-See optimization

# Chapter 1

# Introduction

In the discipline of operations research, optimization tools and other quantitative algorithms are added to support the decision-making process in a highly uncertain environment. The most straightforward approach is through deterministic methods that propose a perfectly-known future. Uncertainty does not exist, and the planning step gives a set of decisions that do not require updates because no new information is introduced. This approach may approximate some uncertainty cases where the effect is minor or when a small variability exists. For other issues, the simplification results in the wrong decisions being made at a high cost, considering that the variability in the system's state increases the complexity and size of the problem. In some cases, the probability distribution or the stochastic process was undefined. If the probability distribution is available, the stochastic approach that treats the continuous distribution as a discrete one (both represent the same probability space with a controlled loss) can be used. Different techniques could be used to solve the problem.

Stochastic processes could lead to different paths, with potential growth in the number of feasible scenarios if the time horizon increases. If the uncertainty realizations are exhaustively enumerated, high computational efforts will be required, resulting in an increase in the solving time in terms of the practitioner's requirements. To face this cardinality, reduction and generation scenario algorithms have been proposed. Questions were derived from investigating the reliability of the forecasts for use in the age of the scenarios, the method of scenario selection (if available) to capture the system's dynamic, and the necessary number of scenarios. Other issues include the probability space where the stochastic process lives because it is not always available or defined; in such cases, different approaches such as the robust and fuzzy optimizations should be used. If the space is limited into several divisions during discretization, the number of possible states or nodes and the size of the problem increase.

Planning tasks is difficult when variability is present, but choosing the tools may make finding a solution harder instead of facilitating it. Planning outcome is a set of decisions linked to the state system; thus, the stability and granularity (number of them) are essential from a practical standpoint. The trade-off between precision and pragmatism is critical to practitioners, contributing to expanding the operations research field.

A consistent way to adjust the system behavior when the original program is affected by a strange event is to have extra capacity waiting for the instance to act or to react more appropriately. Additional capacity refers to the capacity of an organization that is over-maintained for the sole purpose of using it as a reaction when something goes out of schedule. Extra capacity expresses itself in different ways depending on the business, but in all cases, an additional cost is associated. This extra cost could be an anecdote if the unplanned event occurs in such a way that the use of the excess capacity balances the opposing perspective. If the additional capacity is used less frequently, the costs are reduced. The idea of extra capacity is linked to a more general concept, ”flexibility.” Flexibility is typically used so that the definitions could differ slightly. Still, it is generally accepted that flexibility is related to the ability to re-allocate or re-distribute resources (products, capacities, or services) most effectively after any uncertainty (demand, supply, or others) has been revealed (Chen et al., 2018). Its nature is linked to uncertainty management because it gives the possibility to adjust resources depending on the state of the system.

Another aspect to consider is the hierarchical nature of decisions. Most operational decisions are discussed frequently, in some cases, daily; however, tactical decisions belong to the midterm but influence the flexibility or ability to react to variable environments. For example, choosing a particular technology could increase capacity and affect decision-making during strategic planning. In such cases, extra capacity could be acquired and made permanently available, and the idle fraction taken as a sunk cost.

However, not only can flexibility be implemented on resources, but it can also be implemented in the decision-making process, i.e., varying the number of stages or instants where decisions are made, or by anticipating or postponing decisions (Mandelbaum and Buzacott, 1990). "Adjusting" or modifying the originally planned decisions as the different states of nature affects the planning process because the change in the time for making decisions results in additional costs; thus, an economic analysis of this stochastic environment should be conducted.

Agricultural planning is highly affected by this complexity, especially given the number of sources of uncertainty that are present. The market structure, biology of the products, and the management of industrial and farm operations are highly stochastic. Chile's agriculture sector directly accounts for 3% of the country's GDP and approximately 13% of the GDP if relative industries (with high impact in different regions) that create opportunities for local economies were considered (Banco Central de Chile, n.d.). The effective management of this issue in the Chilean economy critically impacts the country's position in international markets and, more importantly, the stability of the local economy. Among other valuable products, Chilean wine stands out.

The *Vitis Vinifera*, the scientific name of the wine grape, is strongly affected by specific geographical and climatic conditions that generate different characteristics of flavor and intensity, which causes variations in the wines obtained. These characteristics must be preserved during industrialization, especially if the taste, flavors, and aspects are factors used in differentiation. However, farm operations are more critical because the input quality for the industrial process has rarely been improved. There are several operations on the farm, and among them, pruning and monitoring the vineyard during the pre-harvest and harvest phases are the most important. The harvest is also of great importance; once the fruit is harvested, the biological cycle in the plant stops, and the degradation process begins. Thus, rapid industrialization is required.

The ripening process can be negatively affected by the effect of rains. There is no consensus on how the intensity of rainfall affects the ripening process since some conditions do not allow the isolation of this influence. However, there is a consensus that indicates that beyond a certain threshold of rain intensity, there is a water absorption effect that brings about a series of inconveniences: low levels of sugar concentration due to dilution effect of the absorbed water, weight gain that can end in the shelling of the bunch, and swelling of the grape grain, which can lead to the breakage of the skin and the consequent development of diseases (Coombe, 1992). For the harvest problem, the decisions to be made are the number of the labor force, their qualifications, the time of hiring and termination, where to harvest, and in what quantity. In this study, we treated the harvesting task for the Vitis Vinifera as a multi-period stochastic problem to understand how this model adds value to the farmer in different conditions of uncertainty and resource flexibility. To do that, we used a simplified (but accurate) decision model as a basis. Three main focus points were developed for this document: 1) generation of a multi-period stochastic model value for decision-making, which was compared to the simplified approach under different conditions of uncertainty and flexible resources, 2) the different times of decision-making and their impact on the value of the solution, incorporating different degrees of flexibility in the environments where the constitution of work teams is feasible, and 3) the size of the problem for multi-period optimization, reported in the literature as highly demanding computationally.

The primary motivation to tackle this problem from a stochastic point of view is the scarcity of the literature for this type of application, with the real problem requiring fine-tuning of the model due to the complexity of the task when several lots are being managed.

To do this, we modeled the harvesting stage using different alternatives, ranging from deterministic to stochastic optimizations, under other conditions of uncertainty (represented by probability distribution) and in the presence of flexible resources of various degrees. In addition to the traditional methods of flexibility (resources), we explored the impact of the number and periods of stages on the expected results of the system. We attempted to understand which conditions are sources of flexibility for the decision-making models. Finally, we determined the program's size, a practical problem linked to the computational effort and complexity.

## 1.1. Objectives

The main objective of this study is to model the decision-making process for wine harvesting under uncertain conditions and solve the different optimization problems that occur considering the existence of flexibility sources; to determine the impact of value generation in an agriculture case under uncertain conditions.

The specific objectives are:

i To study the current literature and synthesize it in a state-of-the-art analysis

ii To model and solve the harvesting problem using the expected value problem approach and multistage stochastic programming variants, considering quality degradation in an uncertain context.

iii To determine the impact on the system's value on different types of optimization models for other conditions of uncertainty when flexible resources are present.

iv To develop, model, solve and analyze a novel approach for reducing the size of the multistage problem.

The contributions of this thesis are threefold:

i Regarding the modeling of the harvesting case in a stochastic way: the stochastic nature of the problem is acknowledged in this thesis, and a model provided that can be solved without simplifying the modeling process. As most studies in the literature use deterministic approaches or reduce the problem to an expected value, the gap between the actual application and the published ones is filled in this.

ii Regarding the impacts of the different sources of flexibility: the available literature states that flexibility is multidimensional. Its value has been studied in the industrial environment, but this kind of analysis is still new in the agricultural field. Even though recourse actions are a natural source of flexibility for a stochastic approach, this work investigates the specific relationship between the different sources of flexibility, uncertainty events, and models that support the decision-making process. This analysis has been recognized as a poorly explored field, and this thesis contributes to that line.

iii Regarding the tractability problem for the multistage stochastic approach: reducing and generating algorithms are the two main methods used to tackle the tractability problem in a multistage stochastic process. In this study, we proposed a novel application to reduce the computational effort for the rolling horizon algorithm, maintaining acceptable performances losses. The proposed model (Rolling Tree Approach) offers the granularity of the decisions in the short term, combined with a simplified view of the future. This avoids the myopic approximations of the other methods in the literature and limits the loss of performance due to the value of the temporal parameters. Additionally, we went beyond the content in the literature to indicate the conditions of uncertainty and the convenience of the novel algorithm.

## 1.2. Wine Industry in Chile

The fig 1.1 shows the annual wine production by country. Traditional winemakers like France, Italy, and Spain contribute to the total output. Aside from Chile, Australia and Argentina are the other new players on the international market.

However, although Chile's production is smaller than other countries, it plays a critical role in international exportation. Table A.1 shows that Chile occupies one of the top four positions at the global level with 9% of the total exported volume. Regarding the commercial impact, wine accounts for 6% of the exports without considering copper and approximately 16% of the agriculture-relative exports (Banco Central de Chile, n.d.).

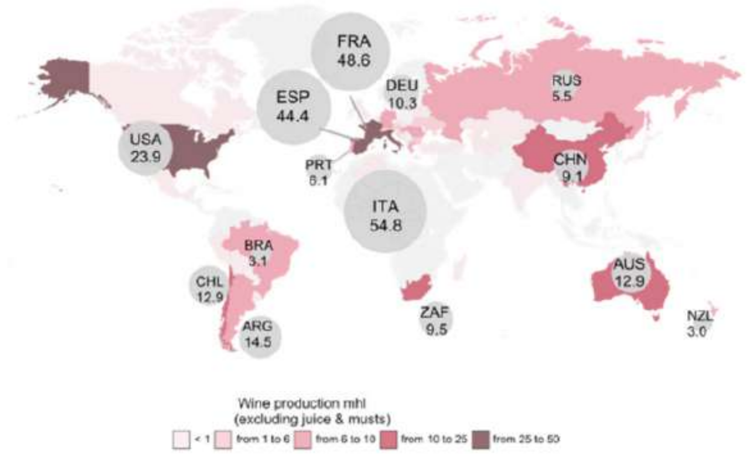
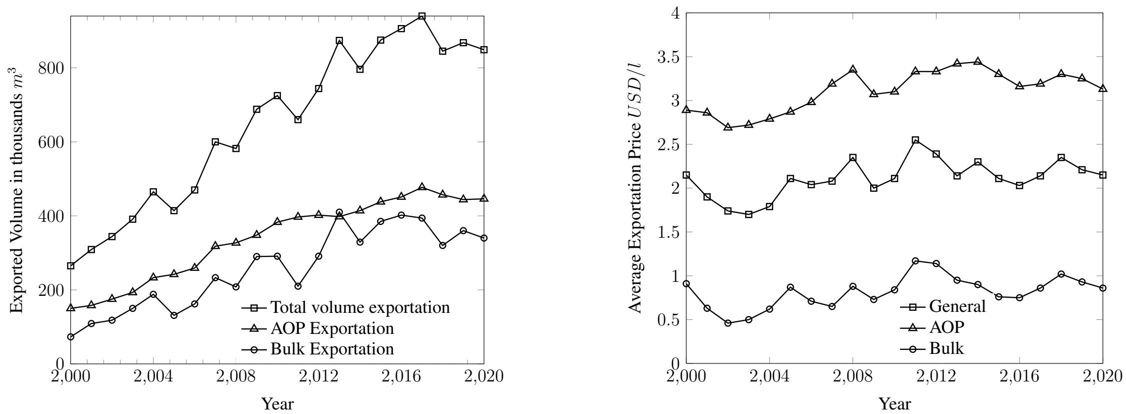



Figure 1.1. Wine Production in 2018 in different countries. Source: International Organisation of Vine and Wine (n.d.)

Table 1.1. Main exporters in millions of hl. Source: International Organisation of Vine and Wine (n.d.)

| **Country** | **2014** | **2015** | **2016** | **2017** | **2018** |
|---|---|---|---|---|---|
| Spain | 22.0% | 23.0% | 22.0% | 21.0% | 20.0% |
| Italy | 20.0% | 19.0% | 20.0% | 20.0% | 18.0% |
| France | 14.0% | 13.0% | 14.0% | 14.0% | 13.0% |
| Chile | 8.0% | 8.0% | 9.0% | 9.0% | 9.0% |
| Australia | 7.0% | 7.0% | 7.0% | 7.0% | 8.0% |
| South Africa | 4.0% | 4.0% | 4.0% | 4.0% | 5.0% |
| Germany | 4.0% | 4.0% | 3.0% | 4.0% | 3.0% |
| USA | 4.0% | 4.0% | 4.0% | 3.0% | 3.0% |
| Portugal | 3.0% | 3.0% | 3.0% | 3.0% | 3.0% |
| Argentina | 3.0% | 3.0% | 3.0% | 2.0% | 3.0% |
| New Zealand | 2.0% | 2.0% | 2.0% | 2.0% | 2.0% |

Thus, this sector is essential and needs its value to be maintained all over the supply chain, which is linked very strongly to the perceived quality of the product.




(a) (a)Annual Exported Volume

(b) (b)Average Exportation Price

Figure 1.2. Exportation statistics: Volume and Price

The figure 1.2shows the evolution of exported volumes for the Chilean case. Two different groups exist based on the exportation data, which depends on the appellation of origin (AO) certificate. Technically, the wines with an AO are from specific regions established by a ministerial decree that reaches certain standards. The total volume of the Chilean exportations has been increasing in the last twenty years, with a slight drop in the previous three years. The AO and bulk exportations follow the same patterns; thus, it is essential to highlight that the AO volume is approximately 15% bigger than the bulk because AOs have a higher added value at the time of export, the value that is created in Chile. The figure 1.2 shows the average exportation price for the 2000-2020 period. It can be seen that the prices are stable, and the AO price is up to 6 times the price of the bulk exports, but the costs are also higher.

Lima, J.L. (2015), China is the leading exporter with 15% of the annual volume for AO wine, followed by Japan, the United Kingdom, and Brazil, with approximately 11% each. These four countries account for about 50% of the monetary terms.

Besides the international market, *Memoria Anual 2019 Vinos de Chile* (n.d.) indicate that the annual inner wine consumption per capita is 14.1 liters (2019). The local market size reached almost USD 1,000 million, with an annual volume of 248.4 million liters. The wine industry accounts for approximately 0.5% of the Chilean GDP and is responsible for 100,000 direct jobs, 53% of who work in vineyards *Memoria Anual 2019 Vinos de Chile* (n.d.). According to *Castastro Viticola Nacional 2019 Vinos de Chile* (n.d.), the surface area of wine grapes reached the 136,000 ha mark in 2019, with 26% of this being white strains. The surface growth has been sustained over time, almost tripling in 1995. This investment requires at least 4-5 years to producesss returns at full capacity Lima, J.L. (2015). A method that reports considerably better yields at an additional cost is the irrigation system, which creates the opportunity to explore new geographical zones of production. The initial costs for the plantations represent 20% of the total cost of the vineyard's life, estimated for 20 years for the purpose of accounting, but with an actual life span of approximately 30-40 years. The price of bottled wine differs considerably from the price paid to producers. Part of this gap is explained by the supply chain of the wine. There are 4 clearly defined participants: grape producer, collectors/intermediaries, wine producers, and wine marketers. This form of commercialization in the Chilean market includes three different ways of contracting:

(i) Long Term: they are expected when the arable land is of good quality for super-premium wines. The winemaking company establishes a relationship with the producer, where the interference extends to agricultural practices. However, this type of contract requires permanent monitoring since the prices paid per kilogram for grape are high compared to other options.

(ii) Annual: in this contract, a quantity range in tons is agreed upon, and a base price is set, which will be corrected by quality parameters.

(iii) Spot Market: there is no contract between the producer and the wine producer before harvesting. In such cases, the intermediaries play the role of negotiators, agreeing on prices based on the demands and conditions of the grape to be purchased.

The contractual strategies should aim to maintain the quality of the wine and at a low cost for wine producers. Typically, a portfolio of contracts, including the spot market volume, is kept to adjust some engaged volume variables or take advantage of overproduction opportunities.

### 1.3. *Vitis Vinifera*

The quality of Vitis vinifera is critical to obtaining a qualified wine. The planning process includes several cyclic steps, such as the annual tasks and others that are fundamental, like the soil choice, which could only be done once. The potential quality of the grape is influenced by the soil selection, weather, and operations during the plant's shelf life. The soil and farm operations are decisions that can be controlled, which diminishes the uncertainty. Still, the weather is essentially uncertain and has been unstable through the years, reducing the effect of its weight which should be high for quality and yield production.

All the efforts at the farm level end during the harvesting stage, a step that will be addressed in greater detail. Here, the process involves determining the degree of maturity, the removal of clusters, and the transportation to the industry. The quality of the harvested material was the maximum that the wine could achieve. Thus, the time and status of the grape should be chosen carefully. Harvesting can be done manually, through mechanized means, or dually. The decision depends on several issues such as the variety, winery reception program, weather, capacity of the vineyard, and availability of labor. Manual harvesting is versatile in terms of the vineyard driving system, but there are ergonomic limitations. In mechanized harvesting, the driving system is limiting, and it is necessary to properly plan the crop development with this orientation. It is advantageous in that operations can be carried out 24 hours a day, which gives it more sensitivity to harvesting during low temperatures to preserve the quality. The flexibility that each option adds to the system is variable. For large extensions, mechanized harvesting could be preferable, even when the machine's capacity is equivalent to between 40 and 50 hand pickers. The cost is less than 20% of the comparable manual labor cost.

In Chile, the land extension is smaller than in wine-grape-producing countries like Spain, and the topography is more irregular. Chile produces different types of wine because of the altitudes and latitudes of the plantations. By local industry information, the harvesting process is done manually for at least 60% of the total product. This is a practical way of managing little farms because it creates the possibility of negotiating the labor cost, thus reducing the fixed costs. Since grapes need permanent dedication throughout the year, and not only during the harvest process, a portion of the workforce was also permanently employed. The harvest can last three months, but the actual demand should be balanced with other factors, i.e., the progress of the ripening process, the forecasting capacity, winery space, and, finally, workforce availability. This situation makes it challenging to ensure consistency of the decisions, so a decision model that supports this process could prove very useful.

### 1.4. Quality and Ripening process

The physiological maturity of wine grape is given by the appearance of the germination capacity; for industrial areas, the selected maturity is technological or related to functionality in winemaking.

The quality of the *Vinis vinifera* is closely linked to the production area (climate and soil), effective average temperature, water regime, rainfall during or close to harvest, and topography of the land. The volatile organic compounds play a critical role in the final quality of the wine; however, the balance in the environment, vineyard practices, and genotypes are poorly understood (Lund and Bohlmann, 2006). This balance must be respected when the harvest time is chosen, considering the acidity, sweetness, taste, and phenolic ripeness. The quality of the grape at the moment of harvest is indicated as the main factor for the quality of the wine (Coombe, 1992). It is difficult to define the optimal maturity because there is no privileged element in the chemical profile to set it (Meléndez et al., 2013). For example, sugar content and acidity are two commonly used parameters, but the ratio changes depending on the type of wine and the winemaker's objective. At least two types of maturities are recognized: technological and phenolic. The first maturity occurred when the sugar content was very high, causing the maturity index (sugar/acidity) to be high. This indicator is typically used in industry contracts for standard grapes; with the acidity being measured using $pH$ (with a scale of $[1, 14]$, where $pH = 1.0$ is extremely acid, $pH = 6.0$ is neutral, and $pH = 14.0$, extremely alkaline) and sugar content through $^{\circ}Brix$. The sugar content was transformed into probable alcohol, equivalent to 16.83 g of sugar/l per $1\%$ alcohol(Glories et al., 2000); p, owing its importance.

Phenolic maturity is reached when the anthocyanin compound concentration is maximum and the tannins content is low for the skin and seed (Le Moigne et al., 2008). Another indicator of maturity is the status of the skin of the grain. Considering other fruits, as the ripening process goes on, the skin becomes softer. In practical terms, it is necessary to incorporate a mix of various ”well-documented” ways of ripening evaluation, especially for products that are designed for higher prices than the standard (Coombe, 1992).

The beginning of the ripening stage is known as *veraison*, the time when the color of the grape grain first changes. This moment varies over the years and is responsible for most of the changes visible on the final day of harvest. The ripening process is developed differently for each berry, so the uniformity is not assured. To control the progress at this stage, a chemical analysis of the juice in the grains was developed. The sugar content increased until the plateau was reached, and the changes after this milestone are explained by the loss of water or the gain of water, but not by the changes in the sugar content (Coombe, 1992).

The maturation stage is highly critical for the development of this raw material. We detailed some changes below:

- Increase in size and weight
- Increase in sugar content, usually to levels close to 200 gr/liter of grape juice
- Decrease in the concentration of acid content
- Color changes towards the specific pigmentation of the vine type

- Increase in the concentration of the aromatic and taste substances, responsible for the organoleptic profile. The observation of the maturation process using these parameters is called phenolic maturation

Tepferrer2008optimization, the harvest should be done at the right moment. Otherwise, value is diminished. Premium wine can be degraded if the harvest time differs from the optimal day. In that study, the authors represented this behavior with a quality loss function based on the Taguchi model (see Besterfield (2003)), introducing the novel model. The proposed curve looked like a parabola and was made using enologist surveys which described the changes following a professional standpoint. In the vinification process, sugar changes to alcohol in a controlled fermentation process. The alcohol content and the high acidity (low pH) allow the wine to be preserved from unwanted microbiological events. Still, they are also part of the desired character of the product.

To go deeper into the wine market, wine grape nature, and vinification process, please refer to the appendix. A

### 1.5. Structure of the thesis

The rest of the document is divided into five chapters. At the beginning of each chapter, there is a vocabulary to be used. The content presented in the different branches is as follows:

**Chapter II** summarizes the research done in the field of agriculture planning, where operations research tools have made a remarkable impact. We also reviewed the research on the wine industry. The use of multi-period stochastic models in agriculture is presented, and we explore the tractability problem in general applications. Reduction and generation methodologies are then briefly explained. Finally, we introduced a short review of the metrics of multistage trees.

**Chapter III** is based on an already published paper by the author of this thesis, which discusses the impact of stochastic models in the valuation of the system for several uncertain conditions and flexible resources. We expanded the quality degradation formula Ferrer et al. (2008) and introduced stochastic modeling for the uncertainty event.

**Chapter IV** presents the different ways through which flexibility determines the interaction and impact of the system value. Two main flexibilities are discussed: epochs of decisions in multistage stochastic models considering the uncertainty realization as a reference and the availability of multiple flexible resource types to perform the task. We determined these flexibilities and performed a sensitivity analysis since the costs are intrinsically related to flexibility. We expand the metrics, including *nested distance*, to measure how different two stochastic solutions are.

**Chapter V** offers a new approach to solve the high computational cost in multistage stochastic problem growths. We discuss an algorithm based on rolling horizon type that concentrates the calculation efforts in the short-term, maintaining the exhaustive tree structure and diminishing the computational charge through simplifications for future periods. We also determine the impact of this approach (Rolling Tree Approach) on the system's value for different conditions of uncertainty and configurations. We also study the decisions for a set behavior and the expected computational effort.

**Chapter VI** presents the main takeaway of this thesis and suggested future research directions.

# Chapter 2

# State of the Art

In this chapter, we give some details on the state of the literature. Initially, a discussion on production planning, and its application in the agriculture sector, is presented. The variability effects, especially the impact of weather, are then described. After that, two different approximations to deal with the variability problem are given: stochastic optimization and flexibility. Additionally, we introduced the performance metrics used to assess the contribution of the approximations. To finish this chapter, we indicated the assistance of this thesis based on gaps in the literature.

**Nomenclature**

$\xi$, uncertain specific event

$\Xi$, uncertainty set of events

$t \in T$, a period of the time span of interest

$\xi_t$, realization of the uncertain event at time $.t$

$\omega$, individual scenario

$\Omega$, collection of the possible scenarios for the problem

$\mathbf{x}$, decisions to be made in a context of uncertainty,

$\boldsymbol{\xi}$, uncertainty matrix,

$f(\mathbf{x}, \boldsymbol{\xi})$, objective function,

$\mathbb{E}_{\xi}$, expected value of the function under uncertainty

$\bar{\xi}$, the expected value of the uncertainty possible realizations

$w \in , \mathcal{W}$, one specific resource in the original set

$w' \in \mathcal{W}'$, one specific resource in the reduced set for the stage $t$.

$\bar{x}(\bar{\xi}, t, w)$, deterministic solution where $w$ is the resource type or object of decision, at time $t$ considering the expected value of the uncertainty possible realizations, $\bar{\xi}$

$\hat{x}$, specific solution for an optimization problem

$RP^t$, value of the complete or original recourse problem,

$EEV^{t,\tau}(x^{1,t-1}, \xi^{t-1})$, value of the simplified model applied over the whole original tree,

$\tau$, time index for the simplified span time,

$VSS^{t,T'}$, value is the difference between the best of the simplification value and recourse problem.

$MSV^{t,T'}$, marginal stage value

## 2.1. Uncertainty in Agriculture Planning

The agricultural supply chain is a highly uncertain environment because of the nature of the raw materials and the supply chain players Ahumada and Villalobos (2009). The biological nature implies the need for additional variability because the product keeps changing over time, affecting the shelf life. However, this erratic behavior is also present in the early stages of the farm's operations. Weather factors, such as rain and temperature, are important sources of uncertainty in agricultural systems. As Ahumada and Villalobos (2009) points out in their review, what differentiates agricultural supply chains from other supply chains is the importance played by factors such as food quality and safety and weather-related variabilities. In the case of wine grape production (see Jones et al. (2005); van Leeuwen and Darriet (2016); Ramos et al. (2008)), weather factors affect the intensity and length of fruit cycles. Moreover, weather conditions directly affect harvesting operations. For example, in some cases, mechanical harvesting methods cannot be used during rainy weather condition due to soil conditions. The work by Ferrer et al. (2008) presents a deterministic optimization model for grape harvesting operations, weighing a quality loss function and bounded time for the tasks (see 3.1.3 for a deeper discussion). Arnaout and Maatouk (2010) developed a heuristic to solve a larger version oishe previous model. However, in both cases, they did not include stochastic elements. Allen and Schuster (2004) studied the harvesting of concord grapes considering "an uncertain length of the harvest season and an uncertain crop size" through a risk management model. Seyoum-Tegegn and Chan (2013) analyzed vineyard investments, considering the yield and uncertainty of the price, and applying real-world options valuation. As Soto-Silva et al. (2016) mentioned, even though the weather and stochastic nature of the product are critical in agriculture, the number of studies that consider this approach are few.

Quality is affected by the weather and hence, the economic value of the product. Quality plays a vital role in the agricultural supply chain since the product is alive and evolves from the moment it is harvested (Van Der Vorst et al., 2009). There have been continuous efforts in previous studies to model and capture the quality degradation of agricultural products. In work by Rong et al. (2011), the authors model the quality degradation of products by time and temperature as they pass through the supply chain in different facilities and transportation modes. Ferrer et al. (2008) Presented a mixed-integer optimization model that considers the cost of harvesting activities and the loss in quality of the grapes due to delayed harvesting—Ahumada and Villalobos (2011) modeled the harvesting, packing, and distribution of crops to maximize revenue. Their model accounted for labor availability, price dynamics, and the variable effects in the product quality due to weather and plant biology through different functions and approximations. The recent work Jonkman et al. (2019) models the perishability of the product in the form of discrete quality categories or the condition of shelf-life limitations.

To face this uncertainty, flexibility is the most common approach, even when operations research tools are used. In our case, there is a history of applications in operations research, but stochastic vision is scarce. Next, we reviewed the OR approach and its flexibility as a tool in an agricultural context.

### 2.2. Operation Research Approximation

Production planning under uncertainty in agricultural systems has been previously studied using different operations research techniques PBorodin et al. (2016) presented in their review. The most common techniques used to handle uncertainty and assist the decision-making process are stochastic optimization, robust optimization, and simulation-based programming. Stochastic optimization has become an increasingly popular tool to model uncertainty in agricultural supply chains (Esteso et al., 2018). Borodin et al. (2016) Points out that most of the stochastic programming approaches are formulated based on: (I) chance-constrained programming (or probabilistic) problems (CCP), and (II) stochastic programming problems without recourse (SP) or with recourse (a type of two-stage programming, TSP).

As examples of optimization in production planning, we found that the work by Bohle et al. (2010), which uses a robust optimization approach for wine grape harvesting, is subject to several uncertainties such as the actual productivity achieved during harvesting. Moghaddam and DePuy (2011) Use a stochastic optimization model with chance constraints to determine the optimal number of acres of hay that a farm should harvest for animal consumption and how much hay to purchase and sell to maximize the total profits of the farm. PBorodin et al. (2014) presented a stochastic optimization model for scheduling a cereal crop harvest at the optimum maturity, using meteorological conditions as a deciding factor. Finally, it Kennedy (2012) examined the applications of dynamic and stochastic dynamic programming on agriculture and natural resources.

According to Behzadi et al. (2018) there is a need for multi-period planning under uncertainty in the context of agricultural systems and supply chains. Some examples of the two-stage stochastic proing approach are (Ahumada et al., 2012; Cholette, 2009; Huh and Lall, 2013; Wiedenmann and Geldermann, 2015), but multistage stochastic programming approach (MS) has not been extensively used in agricultural systems. Guan and Philpott (2011) applied the multistage approach to a production planning problem for Fonterra, a leading company in the New Zealand dairy industry, taking into account the uncertain milk supply, price–demand curves and contracting. Another major application of MSSP is water management for farm irrigation (Q. Li and Hu, 2020; H. Zhang et al., 2017; F. Zhang et al., 2019). More specifically, the studies involving the application of multi-period planning are: Kazemi Zanjani et al. (2011), who investigate a sawmill production planning problem with an uncertainty in the quality of raw materials and demand; Lobos and Vera (2016) who determine the benefits of using a stochastic modelling approach in a sawmill production environment; Veliz et al. (2015) who present a harvesting and road construction decision problem in the forest industry in the presence of uncertainty, modelled as a multistage problem; Chen et al. (2018) who was motivated by a problem in a seed producing company, and finally, Varas et al. (2018) who look at the bottling planning problem for a wine export company facing demand uncertainty.

A multistage stochastic approach could be pertinent for fresh products since their quality changes over time. Designing stochastic models for fresh fruits (e.g., wine grapes) would contribute to the knowledge in the field since such studies are currently scarce, as indicated by Soto-Silva et al. (2016). They also mentioned a need to make flexible decision-making models that help managers make good decisions throughout the food supply chain. This need for flexibility is critical to overcoming the new challenges faced by the agricultural sector, particularly crop production supply chains, which should be reactive and flexible with a high yield at a low cost Borodin et al. (2016). Some of these challenges are detailed by Kusumastuti et al. (2016), who indicated that in harvest modeling and the processing of real-life agri-chain characteristics, there is several productive aspects such as harvesting time window constraints, yield perishability, seasonality, and uncertainties(due to weather conditions or customer demand) that were often not considered in previous studies.

A commonly used algorithm that considers the uncertainty but in a simplified manner is the rolling horizon approach, RHA, S. Sethi and Sorger (1991). In that model, the planning horizon moves forward to update the system's status. Decision policies are only implemented during the immediate periods. Still, for the planning step, the future is accounted for using an expected value and removing the myopic vision in stochastic systems. As was indicated Chand et al. (2002), long-term forecasts lose their effect on the initial decisions. The long-term estimates require better determination because the model is myopic, and the results could change drastically. The rolling horizon is linked to the forecast horizon because it should be used when the last decision is reliable. The stopping-rule depends on the specific problem, but three ideas were cited from Morton (1981): 1) *apparent forecast*: stop when initial decision has stopped fluctuating or the fluctuation is minimum, 2) *near-cost forecast*: when the marginal benefit of increasing the horizon is minimum, the risk is minimized under certain thresholds, and 3) *near-policy forecast*: the initial decision is the most optimal for an accumulative probability of all feasible scenarios. According to the author, a lot of time would be dispensed to estimate the correct horizon, which may be unproductive due to the high cost. However, (Chand et al., 2002) indicated that this is a promising field, because the time horizon has complex results for the different factors, and presents an opportunity for future research. Moreover, the trade-off between extension and decision quality should be addressed as a critical topic through the approximation of evaluations.

Based on this, there are some interesting studies where the RHA was linked to the multi-stage problems, as outlined in the comments below. Tefang1997rolling, the rolling horizon procedure was applied to a flexible scheduling problem. The demand dynamic was captured by the dynamic of the update of the information. There are different triggers to ensure the information is updated and the planning task is renewed, i.e., event-driven or periodic-task. Tealonso2020dealing, they present an application of the scenario trees considering two planning levels, tactical and strategic. The strategic tree is the main level, with each of its nodes containing a set of scenarios that are treated as two-stage problems, and they belong to the tactical group. At the end of the tactical planning period, the scenarios collapsed due to a unique node that is the initial status for planning the strategic tree. This application does not use the idea of a rolling horizon, but it offers a strategy to diminish the natural bushiness of a multistage tree. The cost is that the decisions are divided into two-time horizons, giving less flexibility than the granularity of a unified tree. Devine et al. (2016) It proposed a sequence of stochastic problems to adapt the dynamic approach to a natural gas market. In addition, the rolling horizon algorithm was used to estimate the number of periods to make better current decisions. Finally, the *Value of the rolling horizon, VoRH*, was established to measure the difference between the rolling horizon vision The perfect foresight conditions. In Devine et al. (2014). The authors proposed an optimization problem that updates information periodically, giving the dynamic for a rolling horizon problem. A problem exists in UK's natural gas market, and the stochasticity is in the demand scenarios that are generated using the Monte-Carlo simulation. Updating the information was done daily, and the horizon planning time was five days. Bischi et al. (2019) She proposed a two-phase rolling horizon algorithm to face the energy cogeneration problem. Specifically, the heuristic was initialized using past information, where the rolling horizon was developed weekly, taking into account one week for mixed-integer linear programming, MILP. The process was repeated until the solution converged. Silvente et al. (2018) We applied a rolling horizon approach to a microgrid problem, where the production of renewable energy and the public demand is stochastics. All the fluctuations would make the situation very hard to solve due to its size, even if the problem had been simplified. Two standpoints may be used to cope with the problem: 1) reactive, wherein based on a deterministic solution, the plan was changed according to specific events, and 2) proactive, where different conditions are considered, giving a more conservative policy. In the second approach, stochastic and robust optimizations are used. The rolling horizon approach is mainly a reactive way of scheduling and solves problems iteratively. A similar application is used Champion and Gabriel (2017).

### Tractability Problem: Scenario Reduction and Generation

The computational complexity of MS programming has been reported in (Birge and Louveaux, 2011), (Shapiro and Nemirovski, 2005), (Pflug and Pichler, 2016), among others. The reason is that the number of periods and the possible chain of events increases the final tree size, with the model and computational problem suffering the same destiny. If the original uncertainty set is continuous, it requires a discretization algorithm. This process is part of the area of the generating tree. Generating trees is a general way to refer to the process which leads to a smaller tree beginning with a bigger one. This procedure diminishes the intractability problem. The cost of this approximation is the difference in the final value of the system and the decisions set this t provided. The tolerance to this gap depends on the case, but it should typically be kept short. The trade-off between tractability and gap is linked to specific necessities of the user. The average sample approximation (SAA) (Kleywegt et al., 2002; Kim et al., 2015) and variants (Bertsimas et al., 2018; Pagnoncelli et al., 2009; W. Wang and Ahmed, 2008) were frequently used in the 2000s. However, Shapiro (2010) alerts about the speed-up ramp in the multistage case. Even when the problem is smaller than the original, the size is critical according to more periods. There are several approaches to select the scenarios or the subset of components of the partial tree that will replace the original one to solve it. The reduction path has been revisited many times as an alternative to the discrete approach for a continuous distribution. The reason is that even in a discrete way, the tree could be almost intractable with a high computational effort associated with the solving process. In (Dupačová et al., 2003), the scenario reduction as a procedure for determining the subset of scenarios of a more extensive set was introduced, where both distributions are closer in terms of the probability metrics, including the Kantorovich-related one. In other words, the distance between the original distribution and the proposed one should be small. Dupačová et al. (2003) work with the Forget-Mourier distance metric. This type of technique is called ”moments matching”. More recently, Timonina (2015) recognize the critical role of the approximation methods in solving multistage problem tasks. In Høyland and Wallace (2001) an approach where the generation of scenarios is constrained by a minimization optimization problem is presented, the statistical measure is used as an objective function. This model tries to solve the problem of choosing the scenarios, their number, and how representative they are compared to the original tree. Although this is a versatile approach, the approximated discrete distribution does not always maintain all the values of the distribution moments equal to the original. Hence, the uncertainty in those cases is different in the initial and partial distribution proposals. Inherent to that distribution is the risk analysis, linked to its lower tail. In Høyland and Wallace (2001) the authors allow the user to select the statistical properties to represent the distance, customizing the best for the user.

(Kovacevic and Pichler, 2015) presented a variation of the distance approach, using the nested distance concept presented by Pflug (2010) and recently redeveloped by Pflug and Pichler (2016). This distance extends the Wasserstein distance (see Villani (2003)) for the multistage problem, where the structure is characterized by the filtrations (available information at a specific stage) (see Chapter 4 for more details).

In Timonina (2015), the stochastic process is replaced by a scenario process in a specific probability space, where the distance between the processes is measured by the nested distance defined in Pflug (2010). This study discusses scenario generation and computational efficiency. In Pflug and Pichler (2016), several chapters are dedicated to using the nested distance with a reduction interest while attempting to diminish the business of the original tree. When we opt to use an approximation instead of the original stochastic process, there is a loss. Also, the election of specific scenarios leads to different solution values. The gap between the decisions quantities in the true problem and the approximated are part of the stability analysis. The stability is also a function of the measured values that were used. There exists a complete library about it, see (Heitsch et al., 2006; Heitsch and Römisch, 2009; Ruszczyński and Shapiro, 2003). The main result in (Heitsch et al., 2006) is that the complexity of multistage optimization cases requires the preservation of both filtration and probability distances to achieve high stability.

## 2.3. Flexibility

As it was mentioned in Soto-Silva et al. (2016), there is a need for considering the flexibility types to be assessed.

Flexibility has been recognized in various disciplines as a strategy to manage different uncertainties (Esmaeilikia et al., 2016). There are many definitions of flexibility, and they vary from one discipline/context to another, with confusion surrounding its dimensions and stages (Sawhney, 2006). In the case of manufacturing, flexibility is referred to in various system states that can be adopted to manufacture different product types at other volumes (Slack, 1983; Upton, 1994). That is the ability of a manufacturing system to react to shifts in the various forms of the system with a bit of penalty in time, cost, and performance (Swafford et al., 2006). Manufacturing flexibility and its measures have been well studied in previous research (Beamon, 1999; De Toni and Tonchia, 1998; Koste et al., 2004; Koste and Malhotra, 1999; Swafford et al., 2006).

Flexibility sources in manufacturing environments have received significant attention in recent years; see the reviews (Beach et al., 2000; Chen et al., 2018; Esmaeilikia et al., 2016; Jain et al., 2013; Slack, 1983; Terkaj et al., 2009) as an example of literature related to flexibility in manufacturing. The different sources of flexibility that have been analyzed in the literature have been summarized in Jain et al. (2013)) as: machine, operation, routing, volume, expansion, process, product, production, material handling, program, and market and labor flexibility. For our case, we will focus on: operation, volume, process and labor flexibility. Operational flexibility is defined as the ability of a part to be produced in different ways. Volume flexibility is the ability of the manufacturing system to be operated profitably at different levels of overall output (A. K. Sethi and Sethi, 1990). Process flexibility is defined as the number and variety of products which can be produced without incurring into high transition penalties or large inventory. Finally, labor flexibility can be defined as the number and variety of tasks/operations a worker can execute without incurring into high transition penalties or large changes in performance outcomes (Koste et al., 2004).

Jain et al. (2013) indicate that flexibility increases the responsiveness of the manufacturing processes of a system, as it improves the utilization of the available system resources and enhances the ability of a manufacturing system to cope with internal and external disturbances. They further point out that flexibility is not only a desirable characteristic, but it is quickly becoming a requirement for the survival of production-oriented companies (Arafa and ElMaraghy, 2012; Barad, 2013; Chryssolouris et al., 2013; Patel et al., 2012; Patel, 2011; Shi and Daniels, 2003). In the case of the agricultural sector, there is little research related to production planning optimization models and how flexibility affects the responsiveness of the system—they work by Mezgár et al. (2000) looks at a flexible network of co-operating small and medium size farmers in Hungary, but does not present an optimization model. The recent work by Chen et al. (2018), looks at the value of flexibility in the production of vegetable seeds by upgrading quality after harvest. Some of the authors of this paper have looked at the value of flexibility by analyzing the benefits of postponement in the bottling process in the wine industry (Varas et al., 2018).

In multi-stage stochastic optimization models, the sources of flexibility and its value have not received much attention, especially in the agricultural context. Soto-Silva et al. (2016) indicate that flexible decision support and models play an important role in helping managers through the entire food supply chain, which is in continuous change because of different uncertainties. (Borodin et al., 2016) go further and indicates that to overcome the new challenges facing the agricultural sector, crop production supply chains in particular should be very reactive and flexible, with a high yield at low cost. (Labrianidis, 1995) present a case study, not a mathematical model, of flexibility through subcontracting in the Greek poultry meat industry, indicating their benefits and drawbacks. Lobos and Vera (2016) studied the benefits of using a stochastic modelling approach versus a rolling horizon for the case of a sawmill operation. There is also a number of publications related to multi-stage optimization applied to water resource management (G. H. Huang and Loucks, 2000; W. Li et al., 2010; Zhou et al., 2013).

## 2.4. Valuing Models and Flexibility contributions

### Valuing different models

Different models represent different requirements of information and modeling efforts. To compare, we need to measure the differences. In this section, we visited this topic to understand how to compare stochastic and deterministic approaches.

In (Birge, 1982), the author introduces a way of measuring the contribution of stochastic approaches, considering the expected value. He resumes the *Expected Value of Perfect Information* concept (Madansky, 1959), EVPI, to indicate the necessity of having a useful measure in the cases of incomplete information and a stochastic environment. The *Wait-and-see* solution, WS, is the perfect information environment, where each future realization is known a priori, so the path is solved optimally without nonanticipative constraints among them. He also proposed the *value of stochastic solution* metric, VSS, that looks for the loss in value by comparing a simplified view of uncertainty, i.e., *expected value problem*, EVP, and a solution where stochasticity is taken into account explicitly. In this context, uncertainty derives from a probability distribution, individualized by different realizations denoted by $\xi \in \Xi$, to computationally manage it. In a time span of size $T$, there is a sequence of $\xi$, i.e. $(\xi_1, \xi_2, ...\xi_t, ...\xi_{T-1}, \xi_T)$. Each sequence is called scenario or leaf, represented by $\omega$; their collection is $\Omega$. Considering the minimization problem describes for the function $f$, with $\mathbf{x}$ decisions to be made in a context of uncertainty, $min\ f(\mathbf{x}, \boldsymbol{\xi})$, WS approach is formulated by 2.1, and the recourse problem, RP, problem by 2.2

$$WS: \ \mathbb{E}_{\xi}\left[min_x\ f(\mathbf{x}, \boldsymbol{\xi})\right] \tag{2.1}$$

$$RP: \ min_x\, \mathbb{E}_{\xi}\left[f(\mathbf{x}, \boldsymbol{\xi})\right] \tag{2.2}$$

Measures such as EVPI (Madansky, 1959), VSS for the two stages (Birge and Louveaux, 2011; Birge, 1982) and the extension for multi-period RP (Escudero et al., 2007), provide a base to consider excellent information value, stochastic approach value, and deterministic approximation contribution. This thesis does not use the two-stage approach so the VSS terminology will be used for the MS approximation. The VSS could be computed as the difference of the expected value of the solution for MS and the desired value problem decisions applied over the whole set of scenarios in the RP model. The last approach is called *expected value solution of the expected value problem*, EEV.

In Maggioni and Wallace (2012) the authors go deeper about why the EV decisions could be bad. The question is split into two: decisions/resources/action choosing, and quantities. For 2-stage models, they introduced two new measures, the *loss using the skeleton solution*, LUSS, and the *loss of upgrading the deterministic solution*, LUDS. Both give information about the structure of the solution and not only about VSS. They proposed three tests that basis on Thapalia et al. (2009):

- Test A: $VSS = EEV - RP$, where RP means a stochastic model with recourse, in the case of the authors, the two-stage problem.
- Test B: Being EV model a simpler approach, the decisions for a specific stage are denoted by $\bar{x}(\bar{\xi}, t, w)$, where $w$ is the resource type or object of decision. If $\bar{x}(\bar{\xi}, t, w)$ is in the lower bound, then resource $w$ will not be able to be used in the stochastic approach, replacing it by zero in the stage $t$. For example, suppose that in an assignment problem, there are three types of resources to be used, and one of them is not assigned in the deterministic solution. The not assigned resource is not available for evaluation in the stochastic approach. In that way, we force to use only the resources that were selected from the deterministic point of view. The stochastic optimization problem could be written as an equation2.3. The objective function keeps similar to the original problem, but the available resources to be part of the solution are constrained. The original set of resources, i.e. $w \in \mathcal{W}$ is reduced to those used in the EV problem, so $w' \in \mathcal{W}'$, for the stage $t$.

$$min_{x_{w'}} \mathbb{E}_{\xi} \left[ f(\mathbf{x}, \boldsymbol{\xi}) \right] \tag{2.3}$$

Let $\hat{x}$ be the solution for the problem exposed in the equation2.3. Then, the answer is computed in the original tree structure. The solution's value is called the *Expected skeleton solution value, ESSV*, obtained like an equation2.4. The procedure is similar to the EEV case in general terms.

$$ESSV = \mathbb{E}_{\xi}\left[f(\hat{x}, \boldsymbol{\xi})\right] \tag{2.4}$$

If ESSV=RP, it means that the structure of the solution in the deterministic behavior is similar to the one used in the stochastic solution. This is because the stochastic solution is bounded to the resources selected by the deterministic model. According ESSV and RP gap increases, the deterministic solution is worse. This metric is called *loss using the skeleton solution*, LUSS.

- Test C: Being $\bar{x}_{\xi}$ an expected value solution of the deterministic model, now that solution is used in the stochastic version. The idea is to obtain a matrix of values in the stochastic solution, such that the difference with $\bar{x}_{\xi}$ will say how good is the latter.

  If in the stochastic problem, the expected deterministic solution could be upgradable, then the contribution of the stochastic approach exists. Still, the deterministic solution could be considered a lower bound and is always upgradable. The original problem is reformulated to model2.4. The answer to the problem2.4, $\hat{x}$ is used in the *expected input value, EIV*, calculated by equation2.6. A new metric is introduced, the *loss of upgrading the deterministic solution*, LUDS, $EIV - RP$. If EIV is similar to RP, it means that stochastic decisions are not better than the deterministic one, so then the solution $\hat{x}$ is optimum to the stochastic problem.

$$min_{x \geq \bar{x}_w(\bar{\xi})} \quad \left[\mathbb{E}_{\boldsymbol{\xi}} f(x, \boldsymbol{\xi})\right] \tag{2.5}$$

$$EIV = \mathbb{E}_{\xi}(f(\hat{x}, \boldsymbol{\xi})) \tag{2.6}$$

In Maggioni et al. (2014) they search for bounds in multistage stochastic programming to consider how good the approximations are. They enhance the application of LUDS, LUSS, and VSS in the multiperiod problem. Three different measures groups are presented: measures of information, measures of the quality of the deterministic solution, and the rolling measures which update information. The main contribution is that the measurements are useful for estimating bounds to the approximation value.

An interesting work by Pantuso and Boomsma (2019) on the multistage stochastic recourse problem, MSRP, focuses on the number of stages to be solved. As they claimed, selecting several scenarios could be very difficult when there are a lot. If the horizon is finite, but the number of nodes in each stage increases, the tractability of the problem will be complicated, too, because of its size. In that context, reducing scenarios or the horizon is a way to cope with the problem. The authors in (Pantuso and Boomsma, 2019) try to answer the following questions: "What is the value of solving the original problem compared to the approximation?... What are the costs of approximating the original problem by reducing the number of stages?... What is the benefit of including an additional stage in the approximation? ". In that study, the authors use RHA's rolling horizon algorithm to extend the concept of expected value solution to the MS approach. The objective is to replace gradually the tree structure of the MS problem for the uncertainty expected value, and based on the evolution of the value of the system, the value of an extra stage could be delimited. The procedure uses the idea of the rolling horizon. Considering a problem of $T$-stages, where any intermediate stage is denoted by $T'$, they proposed to split the original time span into two windows, where the first consider the discrete uncertainty set and the second uses the expected value of the uncertainty set. The figure 2.1 appears in their work and is useful for a short example. The problem is a four-stage one in a tree representation. The number of stages is $T = 4$, but the approximation is a two-stage. The two-stage is a very well-defined problem in the literature, and considers the first stage uncertainty and simplified the latter one using its expected values. The $\xi_1$ is known, $\xi_2$ is unknown and periods $\xi_3$ and $\xi_4$ are approximated by their expected value, $\bar{\xi}_3$ and $\bar{\xi}_4$. The first step solution after the optimization is $\bar{x}_0$. In step 2, the solution $\bar{x}_0$ is fixed and $\xi_1$ and

$\xi_2$ are known. Now the expected value of the uncertainty is for stage 4, and we have two subtrees. The procedure follows until the whole original tree is discovered. The quality of this approximation is computed when the solution is used in the original problem. They arrived at the EEV equivalent term in multistage problem, the *Rolling-Horizon Value of the Reference Scenario*, RHVRS.

In Pantuso and Boomsma (2019) as well, trie to extend their research to the bounds determination. They proposed different approximations, for example, the WS model. In that case, the size of the problem and the uncertainty set keeps similar to the MS model, but the NAC relaxation may be an advantage in terms of computational effort. The authors introduce the *marginal stage value*, MSV, to assess the contribution of different stages. First, the value of stochastic solution to the stage-t problem compared to the T'-stage approximation is defined as $VSS^{t,T'} = min_{\tau=1,...,T'}\ (EEV^{t,\tau}(x^{1,t-1}, \xi^{t-1}) - RP^t)$, where the $RP^t$ is the value of the complete or original recourse problem and $EEV^{t,\tau}(x^{1,t-1}, \xi^{t-1}$ is the value of the simplified model applied over the whole original tree. The $\tau$ parameter is the time index for the simplified span time. Finally, as there are different ways of simplifying future events, the $VSS^{t,T'}$ value is the difference between the best of the simplification value and recourse problem. Once $VSS^{t,T'}$ is defined, then marginal stage value (MSV) expression is $MSV^{t,T'} = VSS^{t,T'-1} - VSS^{t,T'}$, the difference of making an approximation of $T'$-stage and $(T' - 1)$-stage. The MSV could be obtained by the difference of the EEV values easily.

They provide a portfolio replication problem as an example. They inform that in their example, according to increasing the number of stages in the approximation, the value of the approximation model performance is more significant too. However, they highlighted that this performance improvement behaves erratically.

Finally, a traditional way to understand the change in the final value of the original tree and its approximation is through sensitivity analysis. Sensitivity was studied Higle and Wallace (2003), indicating the necessity to understand the changes due to the variable form of the coefficients. The post-optimally analysis is used in this work in certain sections

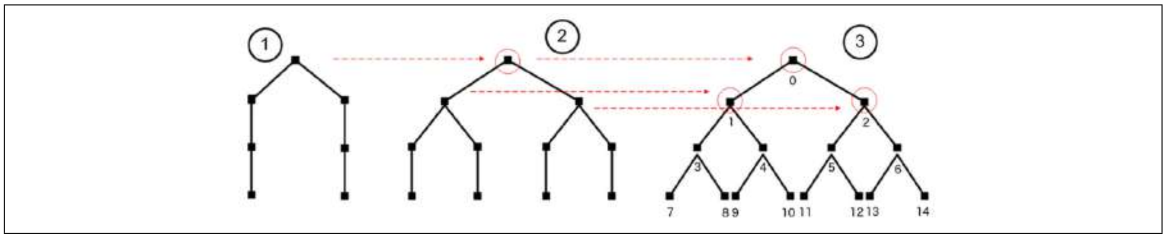



Figure 2.1. The original 'Rolling-horizon two-stage approximations of a four-stage problem. The numbers in circles are the steps of the rolling-horizon procedure. The four-stage scenario tree on the right illustrates the original unfolding of the random parameters, and numbers beneath the node are indices' - source:Pantuso and Boomsma (2019)

to see the effects. Still, it is limited because of the inconvenience that introduces too many parts varying simultaneously. Although statistical analysis gives general regressions that correlate that information, it is not always easy to see in terms of managerial practices.

**Measuring Flexibility**

Measuring the value of accounting for uncertainty and the ability to adjust decisions has not been significantly studied in the literature. In accordance to Rogalski (2011), "Flexibility-related characteristic values are currently not available or defined, which seems to be the crucial difficulty. Reasons for this can, for one, be found in the currently still unresolved problem of universally gauging and evaluating flexibility, which arises from the latter's multi-dimensional character"; in addition, the author indicates that flexibility is usually valued considering that the system gains value because of flexibility very existence, disparaging the environment conditions. K. Huang and Ahmed (2009) propose a very simple way of measuring the input of the decision process: the difference between the values of the objective functions. In their work, they present the case for capacity planning, comparing the values obtained by a multistage stochastic model with a two-stage model. Some analytical methods to measure flexibility can be found in Buzacott and Mandelbaum (2008). In their work, they point out that if we want to measure the value of the added versatility in the system, we can do it by determining the expected value with and without the possibility to alter the decisions, and then take the difference to determine the value of the added flexibility. Kazemi Zanjani et al. (2010) use the differences in optimal objective values to compare a full recourse model, such as the multistage, with the mean-value Deterministic and the two-stage stochastic models for sawmill production planning. Tadei et al. (2019) uses extreme values theory, deriving an asymptotic approximation for the probability distribution of the total utility and determining its expected value; this result is then used by Roohnavazfar et al. (2019) to construct an efficient and accurate approach to estimate the value and the structure of optimal paths in a multi-stage stochastic decision network. Cardin and Hu (2016) determines the value of flexibility by computing the difference in the net present worth between decentralized designs and the real option of expanding the capacity of a waste-to-energy system. We will take a similar approach to what Upton (1994) did, by comparing the expected values.

Measuring flexibility is not simple. Many authors have focused their attention on measuring the effect that each or several dimensions of flexibility have on the organization's performance, such as volume, variety, process, and material handling. There is a body of research related to the empirical flexibility measures, including research related to developing an instrument for measuring and analyzing flexibility (Y. P. Gupta and Somers, 1996), developing models for measurement (D. Gupta, 1993; Jordan and Graves, 1995), use of entity-relationship models to evaluate flexibility (Mishra et al., 2014), developing goodness test for operational measure (D. Gupta and Buzacott, 1996), using transfer functions for measuring flexibility (Alexopoulos et al., 2007; Baykasoğlu, 2009; Buzacott and Mandelbaum, 2008; Das and Caprihan, 2008; Esturilho and Estorilio, 2010; van Hop and Ruengsak, 2005; Kahyaoglu et al., 2002). But there is still not enough significant research in quantitative or analytical flexibility measures, especially in the agricultural sector. One important point when examining the convenience of models is the value that they report to the decision-maker. Hu and Hu (2018) compare the results obtained from a two-stage and a multi-stage programming problem for lot-sizing and scheduling under demand uncertainty, with results indicating that the quality of the decision can improve to more than 10% by using a multi-stage approach.

## 2.5. Summary

In this chapter, state of the art was reviewed. The need to develop more practical tools to face the uncertainty in agricultural planning is evident, mainly because of the high variability of the nature of the product. Weather plays a critical role, and its interaction characterization with the crop is essential in achieving good performance in decisions models. This was the motivation to develop our first declared objective, the description of the system.

To face this uncertainty, the use of flexibility and operations research tools has been suggested in literature reports. Flexible resources are frequently chosen, but their costs are easy to see, but not their value. Operations research tools offer flexibility in decision-making if the uncertainty is tackled to the right degree. But two questions were derived here: 1) how complex should the decision model be to create a real impact in the system? and 2) flexible resources and flexible decision models compete for economic resources, thus, which is more profitable and in what context?. Both questions are addressed through the comparison among different models that consider uncertainty, the epoch of decisions, and flexible resources in Chapters 3 and 4. The first two are how flexibility is expressed in the decisions support model.

Additionally, in the literature, the problem size of the stochastic approach was faced using a reduction algorithm. We think that granularity could be helpful in the short term. In this context, we propose a novel way to meet the problem size. The rolling tree approach presented in the Chapter 5is based on the rolling horizon algorithm but maintains the granularity in the short term. We also discuss the time parameter impacts, something critical in this approximation.

To finish, the literature review considers measuring the gaps between alternatives. We follow the traditional ones, but we use in a novel form the nested distance, a concept that was introduced to reduce trees initially—considering some clues of the Timonina (2015) work, we use this distance to compare similarities between decisions trees. The concept is introduced in Chapter 4 and used in Chapters 4 and 5.

# Chapter 3

# Multistage Stochastic Programming as source of value

In this section, we propose to develop a multistage stochastic optimization model that accounts for the variability in the prevailing climatic conditions and the quality degradation of the product if it is not harvested in the optimal conditions. We extend the previous model proposed by Ferrer et al. (2008), adding the effect of climatic conditions. We assume the daily probability of rain according to a proposed transition matrix (Urdiales et al., 2018), which affects the productivity level of the labor force. We will compare the results obtained using two different levels of recourse actions of our MS model, with those obtained using the expected value of the expected solution (EEV) proposed by Birge and Louveaux (2011) and a wait-and-see (WS) approach. We will compare the objective values using standard metrics, such as Value of Stochastic Solution $VSS$ and Expected Value of Perfect Information $EVPI$, and the way in which that value is created. Also, we will analyze how the quality, level of uncertainty, and the impact of rain on labor productivity affects this value. The contributions of this research are twofold. First, we present a MS approach for grape harvesting, where the quality of the product is degraded and the likelihood of rain is uncertain. Second, we will compare the value of different level of recourse actions of the MS approach to the value of the EEV and WS approach and determine if the ability to update information in the planning stage generates value.

The document is organised as follows. In Section 3.1 we present the problem in mathematical way, the optimisation models, and the quality and rain concepts. Section 4.4 shows the main results, for later discussion and conclusion in Section 3.3.

**Nomenclature**

$\omega \in \Omega$, a specific scenario or leaf of the whole set of scenarios of the tree

$\Omega$, the set of scenarios or leaves.

$\Omega'_t$, the set of scenarios or leaves that present state equal to one at time $t$.

$\Omega_g$, set of scenarios in node $g \in \mathcal{G}$.

$g \in \mathcal{G}$: set of nodes

$\mathcal{G}_t$: set of nodes in stage $t\ :\ (\mathcal{G}_t \subset \mathcal{G})$, $t \in \mathcal{T}$.

$\omega_g \in \Omega_g$: set of scenarios in node $g \in \mathcal{G}$

$\tau_{0,1}$: transition factor between two consecutive rainy periods

$\phi \in [0, 1]$: skill level of labor force

$\hat{\beta}_m$: nominal productivity for the resource $m$. If there is only one type, the sub-index is avoided.

$\beta^{\omega}_{tm}$: effective worker $m \in \mathcal{W}$ productivity at time $t \in \mathcal{T}$ in scenario $\omega \in \Omega$ (kilograms per worker per period).

$\breve{\beta}_{t,m}$: actual deterministic productivity for the resource $m$ at moment $t$

$\xi \in \Xi$: the set of possible values that may take the uncertainty realization. Binary values.

$\bar{\xi}_t$: expected realization in period $t \in \mathcal{T}$.

$\dot{x}$, decorator for variables (i.e. $x$) that are decided before uncertainty of the period is realizated

$h^{\omega}_{jt}$: daily harvested quantity at block $j \in \mathcal{J}$ in period $t \in \mathcal{T}$ in scenario $\omega \in \Omega$, calculated as $\beta^{\omega}_t z^{\omega}_{jt}$ (kilograms/day).

$x^{\omega}_{tm}$: workers $m \in \mathcal{W}$ hired at time $t \in \mathcal{T}$ (workers) for path $\omega \in \Omega$ (when EV problem, $\Omega = \{1\}$, and belongs to the expected value of the uncertainty).

$y_{tm}$: workers $m \in \mathcal{W}$ laid off at time $t \in \mathcal{T}$ (workers) for path $\omega \in \Omega$ (when EV problem, $\Omega = \{1\}$, and belongs to the expected value of the uncertainty).

$\dot{z}_{jtm}$: workers $m \in \mathcal{W}$ allocated in block $j \in \mathcal{J}$ in period $t \in \mathcal{T}$ before uncertainty happens (workers) for path $\omega \in \Omega$ (when EV problem, $\Omega = \{1\}$, and belongs to the expected value of the uncertainty).

$z_{jtm}$: workers $m \in \mathcal{W}$ allocated in block $j \in \mathcal{J}$ in period $t \in \mathcal{T}$ after uncertainty is revealed (workers) for path $\omega \in \Omega$ (when EV problem, $\Omega = \{1\}$, and belongs to the expected value of the uncertainty).

$l^{\omega}_{mt}$: manpower or labor force $m \in \mathcal{W}$ at time $t \in \mathcal{T}$ (workers) for path $\omega \in \Omega$ (when EV problem, $\Omega = \{1\}$, and belongs to the expected value of the uncertainty).

$\mathcal{T}$: set of stages in the time horizon.

$j \in \mathcal{J}$: a specific block $j$ of the set of blocks of the vineyard.

$m \in \mathcal{W}$: a specific manpower resourcer $m$ of the complete set. If there is only one type, the subindex is avoided.

$r_{01}$: probability rain for two consecutives periods when first period is dry and the second is rainy

$r_{11}$: probability rain for two consecutives periods when both periods are rainy

$w^\omega$: conditional probablity for the specific scenario $\omega$

$a_j, b_j, c_j$: quality parameters for the cuadratic equation that represents the quality of the grape in the block $j$

$u_j$: fractional quality loss per rainy period for the grape in block $j$

$B_j$: price of the grape in lot $j$ ($/kilograms).

$C_{E,m}$: cost of hiring ($/worker).

$C_{F,m}$: cost to lay off ($/worker).

$C_{P,m}$: cost of keeping labor idle between periods ($/worker per period).

$C_{H.m}$: cost of harvesting ($/kilograms).

$C_{Z,m}$, cost of assignment before uncertainty is revealed by worker $m \in \mathcal{W}$ ($/kilograms).

$C_{\dot{Z},m}$, cost of assignment after uncertainty is revealed by worker $m \in \mathcal{W}$ ($/kilograms).

$K$: maximum daily reception capacity of the winery (kilograms/day).

$S_j$: initial amount of grapes in lot $j$ (kilograms).

$Q^\omega_{jt}$: daily quality of the wine grape at block $j \in \mathcal{J}$ in period $t \in \mathcal{T}$ in scenario $\omega \in \Omega$.

$\bar{Q}_{jt}$: average quality for block $j \in \mathcal{J}$, $t \in \mathcal{T}$.

$\breve{Q}_{jt}$: actual deterministic quality for the block $j$ at moment $t$

$\mathbb{U}_{\mathcal{M}}$: expected value of the solution of a model, i.e., stochastic one.

$\mathcal{M}$: represents any model ant it is useful to write general expressions

$\mathbb{I}_{\mathcal{M}}$: actual income as a percentage of the maximum feasible income for the model $\mathcal{M}$

$\mathbb{L}_{\mathcal{M}}$: cost of labor as a percentage of the maximum income

$\mathbb{Q}_{\mathcal{M}}$: percentage deviation of grape quality from optimum conditions

$\mathbb{S}_{\mathcal{M}}$: percentage of unharvested grapes

$\xi_t^{\omega}$: the value of the uncertainty realization at moment $t$ for the specific scenario $\omega$

$t \in \mathcal{T}$: specific period time in the time span

Note 1: when there is a unique type of resource available, the sub index $m$ is avoided.

Note 2: when there is a unique scenario (i.e., deterministic model), the supra index $\omega$ is avoided.

### 3.1. Problem Formulation

In this section, we will first present a stylized version of the deterministic model in Ferrer et al. (2008). Second, we will discuss the way we model rain uncertainty. Third, we introduce the quality approach used in our model as an extension of the one in Ferrer et al. (2008). We also explain the way we modeled worker ability and the way it is affected by rain. Finally, we formulate the expected value and the multistage stochastic models, both considering uncertainty and quality effects.

#### 3.1.1. The Grape Harvesting Problem

Our model is based on Ferrer et al. (2008) formulation for a vineyard harvesting planning problem with quality degradation present. The vineyard is divided in little pieces of lands, called *blocks* or *lots*. The criteria for this division is not unique, but it is connected with the product inside, the facilities, and the type of job that require, among others. The prevalent idea in the division is to obtain blocks of lands that require similar types of work, to make easier the job scheduling task. The authors' goal is the cost minimization and quality maximization. For this, they introduce a quality loss function that generates extra costs when harvesting deviates from the ideal date. Harvesting as planned may use several blocks with different initial quality states, making the routing of operations through time part of the problem. The results give a good approach for a practical decision support system, but uncertainty was not considered.

In this work, we assume that a manager needs to plan the harvest of different blocks that contain wine grapes and her/his goal is to maximize the profit. The operational harvest plan is generally performed one week before its execution, hence we will use a 10 day span for harvest planning. During that time, on each day he/she has to make a decision regarding the amount of workers hired and dismissed, and how they are allocated. The allocation process renders the harvesting capacity for that period and blocks and grapes are harvested and sent to a winery the same day, where the daily reception is made according to a bounded capacity of the winery.

The costs are mainly given by labor force. There are costs of hiring, termination, and costs of keeping labor between periods. Additionally, there is the harvesting cost which is a productivity payment. Even when the harvesting cost is more important than the cost of keeping labor, we make this difference because in Chilean context, it is common to have a stable seasonal team that remains as a part of the vineyard.

Income is produced by selling the harvested grapes at a market price. The actual price, however, is affected by the final quality and type of the grapes, which depends on the specific harvest time. We represent this by a quality factor, which is equal to $1$ when the harvesting day is the optimal day, and less than $1$ otherwise (later in this document, we present the mathematical model). We tested two different varieties of wine grape, standard and premium, where the time pass has different effects. As prices are according to the quality, the benefits change; the progress of the harvesting depends on the balance of cost and benefits, so very poor qualities probably will not be harvested depending on the kind of grape.

### 3.1.2. Rainfall Events

Rain events could be characterized both by occurrence and intensity. The occurrence and intensity effect of rain will be modeled using a 0-1 binary event, which is defined by $\xi \in \Xi : \{0, 1\}$ where 1 corresponds to the rain event and 0 to no rain. This might be seen as an oversimplification of the rain effect. However, in the work by Haeger and Storchmann (2006), the authors studied the effect that climate, craftsmanship and critics have on the prices of American Pinot Noir. In their work, they found that rainfall during the ripening period and the harvest plays virtually no effect on the price of grapes and hence the quality. On the contrary, the work by Ashenfelter et al. (1995) indicates that rainfall during the ripening period and the harvest had a negative effect on the rices of Burgundy french wines. Haeger and Storchmann (2006) analyzes this contradicting effects between Burgundy and Oregon wines and conclude that precipitation variables during the ripening period in Oregon have no effect, because it only rains 14 ml compared to Burgundy were rain during that season is significant with 30-year average of 127 ml. This is an indication that the intensity of the rain has an effect on the quality once a certain threshold or level has been achieved, but we see no direct effect between the intensity and the loss of quality. So the use of a 0-1 binary process can adequately represent the rain event and probability of occurrence of the event will represent the intensity, indicating when the rain exceeds the threshold level and affects both the quality of the grapes and also the harvesting process.

The dynamics of rain probabilities have been studied and are reported in weather and atmospheric science. The available information is usually a weather forecast where rain is reported as a probability. The forecast time span is variable and requires frequent updates, leading to planning reviews. We use a convenient model for our multistage programming. The model is a two-stage first-order Markovian chain representation, where the probability of rain on a specific day depends on the rainfall status from the previous day (Richardson and Wright, 1984).

As shown in Figure 3.1, the transition probability between states in two consecutive periods $t-1$ and $t$ is represented by $r_{\xi_{t-1},\xi_t}$. We define the *transition factor*, $\tau_{01}$, as the

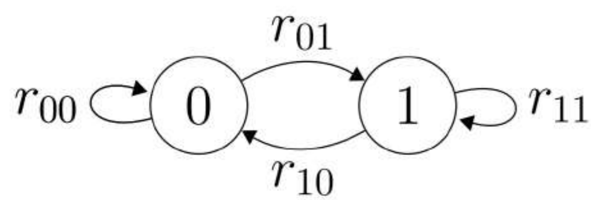



Figure 3.1. Markov Chain for rain transition probabilities. Nodes 0 and 1 represent no rain and rain events, respectively.

quotient $r_{11}/r_{01}$ which represents the propensity of the system to continue in the rain event ($\tau_{01} > 1$) or continue in the no rain event ($\tau_{01} < 1$).

Urdiales et al. (2018) presents a rain model implementation for the Chilean case; they propose an expression which takes into account the "El Niño" climatic event, which generally brings rains in what they define as an *El Niño–Southern Oscillation* (ENSO) index, the geography for several places and different months. The result is a general equation that estimates the probability of rain considering only the rain realization of the immediate period before. According to the information reported, the transition probabilities are stable for each month, year, and location. In our Markov representation we will consider three different transition states: to a smaller rain probability, to a larger rain probability and to one that offers little change (almost constant in practical terms).

According to the Markov representation that we choose for uncertainty, the temporal structure is similar to a tree, specifically a binomial tree. This representation is very convenient and we discuss it later. However, it is important to say that each leaf in the tree is denoted by $\omega$, being $\Omega$ the complete set. The sequence of uncertainty realizations is another way to describe the scenario.

Let $\xi_t^\omega$ denote the value of the uncertainty realization at moment $t$ for the specific scenario $\omega$. Then, for this particular scenario, we can compute the probability of occurrence of the event $w^\omega$, as:

$$w^\omega = \prod_{t \in \mathcal{T}} r_{\xi^\omega_{t-1}, \xi^\omega_t}$$

### 3.1.3. Grape Quality

As Ahumada and Villalobos (2009) points out, since in agriculture we work with living organisms, quality degradation in the supply chain is a differentiating factor from other productive systems. Wines are categorized into quality groups starting at the icon level, premium level, and going down to the reserve, the varietals, and finally the bulk wines (Ferrer et al., 2008). As the actual harvest date of the grapes deviates from the optimal maturity date, the quality of the grapes is affected; a premium wine quality grape could be degraded to a reserve wine quality grape, and if no action is taken it could be finally degraded into a bulk wine quality grape. Hence, during the wine harvest, quality must be carefully managed to obtain the best product at an adequate cost (Coombe, 1992).

During harvest, the quality of the grape is represented by its maturity. However, three types are recognized: technological, phenolic, and chemical maturity (Le Moigne et al., 2008). In this work, we concentrate on the technological maturity, which relates to the degree of sugar content or *Brix degree* and the acidity level, which renders the final alcoholic degree of a wine during the fermentation process. We will focus our attention on the evolution of short-term maturation, which happens during the final weeks, since it drives the operational planning of the harvest. The harvesting task is planned in detail 5 or 6 days before the optimum day to coordinate the resources needed to proceed. During the harvest, the planner must consider the cost of the resources needed to harvest the grapes on an adequate time-frame, and the cost associated with quality degradation due to deviating from the optimal harvest day.

In the work by Ferrer et al. (2008), the authors modeled quality loss as a function of time that reaches the maximum at the optimal day. Before or after the optimum, quality drops. One way of representing this function is by means of a parabola with equation $Q_{jt} = a_j t^2 + b_j t + c_j$, for a specific block $j$ and time $t$, which has a certain variety of

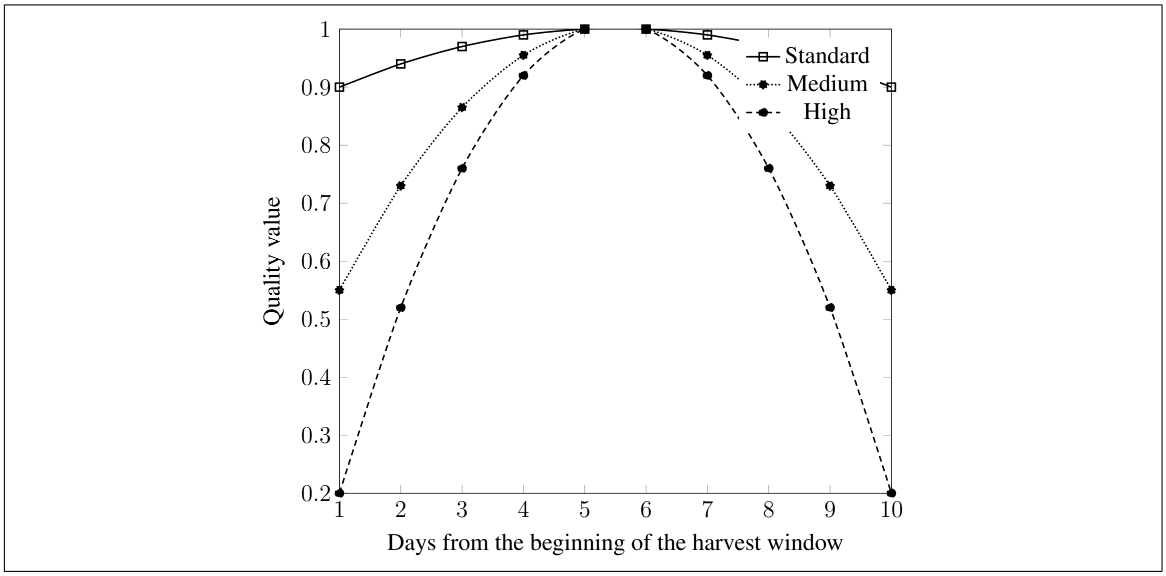


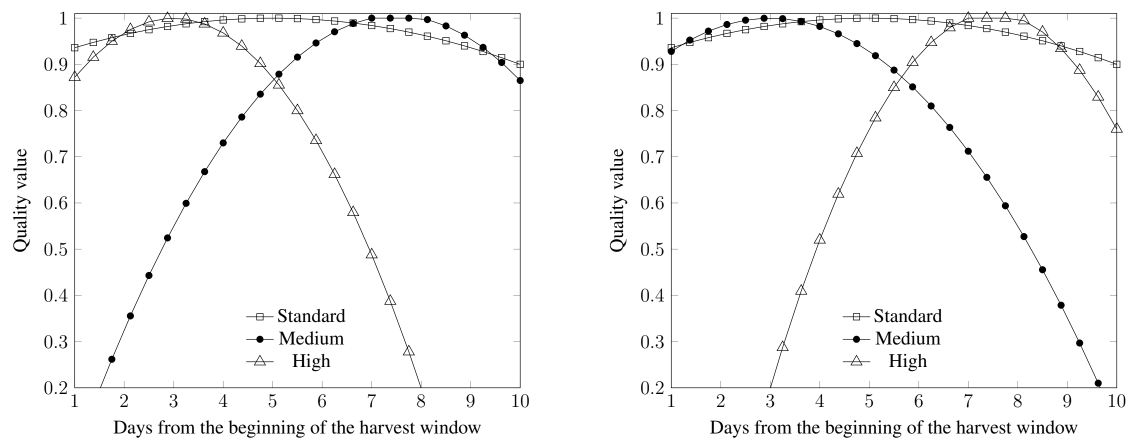

Figure 3.2. Quality degradation curves for three different types of grapes.

grapes. The parameters of this function have to be chosen in such a way that the values $Q_{jt}$ are not negative. Hence, the quality value is in the interval $[0, 1]$. In Figure 3.2, we show three curves corresponding to grapes with three different levels of quality: standard, medium, and high.

About the ripening pattern, we see in Figure 3.3 two extra ripening sequence: 1) medium quality (M), standard quality (S), premium quality (H), or simply, MSH, and 2) HSM. The difference among them is the moment of the optimum maturity.

In this representation, when $Q_{jt} = 1$ the optimum Brix degree has been reached, and it is the right time to harvest. Any change in the Brix degree is represented by a proportionate change in $Q_{jt}$. The three quality curves that we show correspond to different specifications of wine. High quality grapes correspond to a wine that needs very specific features, while a standard grape is used for a wine that offers no differentiation.

In this work, we extend the quality concept to introduce the damage produced by rain. Experts indicate that the absorption of raindrops into the grape's grain decreases the sugar content, so the quality, in terms of alcoholic degree, also decreases Coombe (1992). The




(a) HSM

(b) MSH

Figure 3.3. Ripening pattern for two different qualities of grapes considering the optimum day of maturity. The market price is inversely proportional to the narrowness of the maturity curves.

effect varies according to the type of grape. This process is reversible under the assumption that the grain skin is not damaged, but will result in a harvesting delay because grapes require sunny days with high temperatures "to dry." This usually implies commercial penalties because it impacts industrial planning. In real terms, rainfall bounds the maximum available quality in a specific time window because we cannot wait for the drying process. Therefore, while the potential quality evolution is similar in all scenarios, the achievable quality depends on the particular sequence of rain events in that scenario. The penalty for a sequence of rain events in a specific path, $\omega$, is a function of the uncertainty realization in that scenario between the initial period and $t$, $\xi_t^\omega$.

$$\text{Rain penalty} = 1 - u_j \sum_{\tau=1}^{t} \xi_\tau^\omega \tag{3.1}$$

where $u_j$ denotes the fractional loss of quality per period and has to be calibrated so that the above expression remains in the interval $[0, 1]$. Since the focus of this research is to study the effect that quality degradation (due to rain or deviation from optimal date) has on the harvesting decision, we will keep the optimal harvest date unchanged.

Final quality, then, is given by the following expression:

$$Q_{jt}^{\omega} = (a_j t^2 + b_j t + c_j)(1 - u_j \sum_{\tau=1}^{t} \xi_{\tau}^{\omega}) \tag{3.2}$$

We notice that with this effect quality becomes a random variable affected by rain.

### 3.1.4. Labor Productivity

The effective worker productivity at each day will be defined as $\beta_t$ [1]. The nominal or base productivity of the worker will be $\hat{\beta}$, which corresponds to the harvesting capacity of a single worker in daily base in a dry period. When it rains, the productivity of the worker is affected and so the actual productivity is less than the nominal. We will use a factor $\phi \in [0, 1]$ to account for this reduction in productivity effect due to rain, and we will name it "resource skill" or "ability" of the worker to handle the rain effect. The closer this factor is to 0, the higher the effect the rain has on the productivity and the worker has "lower ability"; as this factor comes closer to 1 the worker is less affected by rain so he/she has a "high ability". The relationship between rain and productivity is given by equation 3.1.4.

$$\beta_t = \hat{\beta} - \hat{\beta}(\xi_t - \xi_t \phi) \tag{3.3}$$

### 3.1.5. Deterministic Model

In the deterministic model, the realization of uncertainty is known at the beginning of the planning. We decorated productivity, $\breve{\beta}$ and quality, $\breve{Q}_{jt}$, to emphasize their deterministic condition :

[1] If there is more than one harvest resource available at time $t$, sub-index $m$ should be added. In the same way, if there is more than one scenario, super-index $\omega$ is added

$$
\begin{aligned}
\max \quad & \sum_{j\in\mathcal{J}}\sum_{t\in\mathcal{T}}((B_j\breve{Q}_{jt} - C_H)h_{jt} - C_E x_t - C_F y_t - C_P l_t) \\
s.t. \quad & \\
& l_t = l_{t-1} + x_t - y_t \quad \forall t\in\mathcal{T} \qquad (d1) \\
& \sum_{j\in\mathcal{J}} z_{jt} \le l_t \quad \forall t\in\mathcal{T} \qquad (d2) \\
& \sum_{j\in\mathcal{J}} \breve{\beta}_t z_{jt} \le K \quad \forall t\in\mathcal{T} \qquad (d3) \\
& h_{jt} \le S_j - \sum_{\tau=1}^{t-1} h_{j\tau} \quad \forall t\in\mathcal{T}, j\in\mathcal{J} \qquad (d4) \\
& x_t, y_t, l_t \ge 0, \in \mathbb{Z}_+ \quad \forall t\in\mathcal{T} \qquad (d5) \\
& z_{jt} \ge 0, \in \mathbb{Z}_+ \quad \forall t\in\mathcal{T}, j\in\mathcal{J} \qquad (d6)
\end{aligned}
$$

In objective function we have the profit, which is given by the price of the grapes multiplied by its quality factor minus the harvest cost, then multiplied by the quantity harvested minus the hiring costs, idle cost and lay off cost. Expression (d1) is the manpower balance, while relation (d2) limits the number of allocated resources. This restriction gives the opportunity to keep workers without going to harvest, if costs are convenient. Relation (d3) bounds the total daily harvest in terms of daily reception capacity of the wine industry. Relation (d4), establishes that the daily harvest is bounded by the remaining volume available in the block. Finally, relations (d5) and (d6) establish the nature of the variables.

### 3.1.6. Model Formulation with Uncertainty

We will now present three different formulations of the problem which account for uncertainty: first, a *Multistage Stochastic Problem with Recourse Actions* (MA), which accounts for uncertainty in rain events and can perform worker re-assignment to blocks after the rain uncertainty reveals itself; second, a *Multistage Stochastic Problem without recourse actions* (MB) formulation, that is quite similar to the MA but does does not allow for worker re-assignment after the rain uncertainty reveals itself; and finally a third formulation, an *Expected Value Problem* (EV) formulation in which uncertainty is reduced to its expected behavior. The three models represent different degrees of flexibility in the decision making process, through the consideration of updating information and active management. Even when MA is the most familiar approach, MB may be required if harvesting blocks are significantly space apart, so its value could be of interesting in practical terms. We will also present the *Wait and See* (WS) approach in order to determine an upper bound to the system.

The notation used in the stochastic models is based mostly on the one proposed by Escudero et al. (2007). In this case the number of periods will refer to the number of the stages, so the index $t$ will be use now as stage in the time horizon.

#### 3.1.6.1. Multistage Stochastic Programming with recourse action (MA)

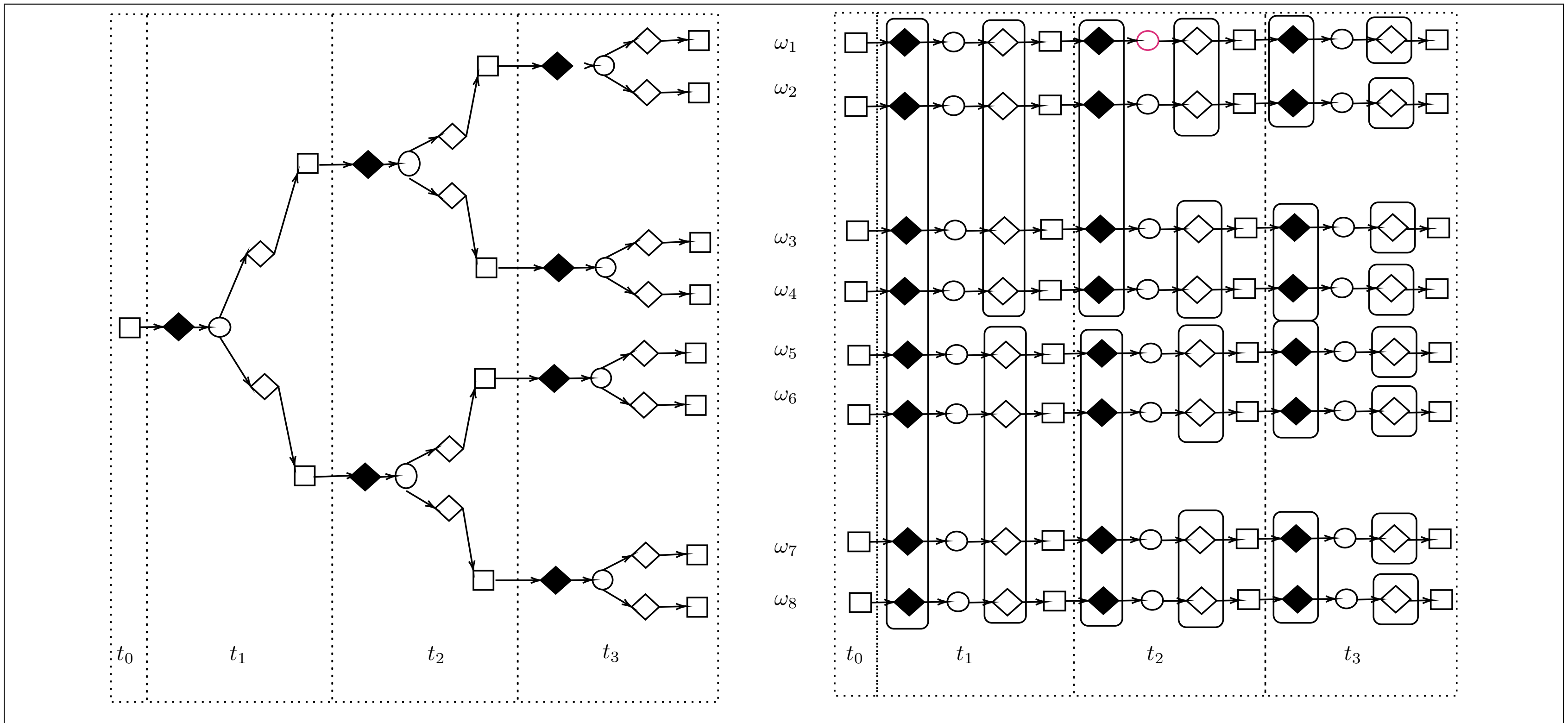
Multistage Stochastic Programming allows to account for all the states of nature, however when accounting for all the scenarios the model can easily become intractable. To avoid the intractability of the model we will use a tree representation of the state space which is commonly used for representing agricultural systems (C. Zhang et al., 2017; Dai and Li, 2013). Also the use of stochastic modeling allows flexibility in terms of anticipating or postponing certain actions (Mandelbaum and Buzacott, 1990). In this context, we will consider what decisions and the moment in which they are taken. In our proposed Multistage Stochastic Programming with recourse action (MA) formulation, the hiring and firing processes are decisions that must be made at each stage before the realization of uncertainty occurs, and the allocation decisions are made after uncertainty is revealed, requiring an active management (when decisions are made **after the uncertainty is revealed**, we decorate the decision variable, i.e. $\dot{x}$). The model's goal is to maximize the expected value of revenue for all the scenarios, keeping the limitations about future information. Because of its linearity and discrete probability distribution, we represent the MA model as an enumeration of the different scenarios in its *deterministic equivalent model* representation (Figure 3.4). To preserve the lack of information in the decision process, the *nonanticipativity principle* is used (Escudero et al., 2007; Mulvey and Ruszczyński, 1995; Rockafellar and Wets, 1991). The MA model is as follows.

$$
\begin{aligned}
\max \quad & \sum_{\omega\in\Omega} w^{\omega} \sum_{j\in\mathcal{J}} \sum_{t\in\mathcal{T}} \{(B_j Q_{jt}^{\omega} - C_H) h_{jt}^{\omega} - C_E x_t^{\omega} - C_F y_t^{\omega} - C_P l_t^{\omega}\} \\
s.t. \quad & \\
& l_t^{\omega} = l_{t-1}^{\omega} + x_t^{\omega} - y_t^{\omega} && \forall t \in \mathcal{T},\, \omega \in \Omega && (s1) \\
& \sum_{j\in\mathcal{J}} \dot{z}_{jt}^{\omega} \le m_t^{\omega} && \forall t \in \mathcal{T},\, \omega \in \Omega && (s2) \\
& \sum_{j\in\mathcal{J}} \beta_t^{\omega} z_{jt}^{\omega} \le K && \forall j \in \mathcal{J}, t \in \mathcal{T},\, \omega \in \Omega && (s3) \\
& \beta_t^{\omega} \dot{z}_{jt}^{\omega} \le S_j - \sum_{t'=1}^{t-1} \beta_{t'}^{\omega} \dot{z}_{jt'}^{\omega} && \forall j \in \mathcal{J}, t \in \mathcal{T},\, \omega \in \Omega && (s4) \\
& x_t^{\omega} = x_t^{\omega'} && \forall\, \omega', \omega \in \Omega_g,\, \omega' \neq \omega, g \in \mathcal{G}_{t-1}, t \in \mathcal{T} \,:\, t \ge 2 && (s5a) \\
& x_1^{\omega} = x_1^{\omega'} && \forall\, \omega', \omega \in \Omega,\, \omega' \neq \omega && (s5b) \\
& y_t^{\omega} = y_t^{\omega'} && \forall\, \omega', \omega \in \Omega_g,\, \omega' \neq \omega, g \in \mathcal{G}_{t-1}, t \in \mathcal{T} \,:\, t \ge 2 && (s6a) \\
& y_1^{\omega} = y_1^{\omega'} && \forall\, \omega', \omega \in \Omega,\, \omega' \neq \omega && (s6b) \\
& \dot{z}_{jt}^{\omega} = \dot{z}_{jt}^{\omega'} && \forall\, \omega', \omega \in \Omega_g,\, \omega' \neq \omega, g \in \mathcal{G}_t, t \in \mathcal{T}, j \in J && (s7) \\
& x_t^{\omega}, y_t^{\omega}, \in \mathbb{Z}_+ && \forall t \in \mathcal{T},\, \omega \in \Omega && (s8) \\
& \dot{z}_{jt}^{\omega}, \in \mathbb{Z}_+ && \forall t \in \mathcal{T},\, j \in J, \omega \in \Omega && (s9)
\end{aligned}
$$

Model 1. Multistage stochastic model with recourse action

Constraints $(s1), (s2), (s3)$ and $(s4)$ are the stochastic reformulation of deterministic constraints $(d1), (d2), (d3)$ and $(d4)$, respectively. The *nonanticipativity principle* is represented by $(s5a), (s5b), (s6a), (s6)$ and $(s7)$ constraints. Constraints $(s8)$ and $(s9)$ are connected to the nature of the variables.

As indicated before, we used a binomial tree to represent the stochastic model (Pflug and Pichler, 2016) as is shown in Figure 3.4 $(a)$. In that tree, the squares represents the nodes or status of the system, a full-filled diamond means a decision made before the realization of uncertainty, and the empty diamonds are decisions made after uncertainty is revealed. Finally, circles are the uncertainty revelation moments. In this work, time and stages are similar.




$(a)$ Scenario Tree

$(b)$ DEM for Scenario Tree

Figure 3.4. Binomial Tree $(a)$ and the Deterministic equivalent representation $(b)$ of the MA model, based on Rockafellar and Wets (1991). References: square:node, full-fill diamond: decision made before uncertainty is revealed, empty diamond: decision made after uncertainty is revealed, circle: uncertainty realization

#### 3.1.6.2. Multistage Stochastic Programming with no recourse action, MB

As it was previously discussed, in some cases the assignment of workers to harvesting blocks needs to be done previously to the event in which the uncertain event has been realized. To account for this we introduce the MB model, in which we do not have the possibility of recourse actions once uncertainty has revealed. This will allow us to determine the value of the recourse actions. We will now present the list the constraints that have been modified in the MB:

$$z^{\omega}_{jt} = z^{\omega'}_{jt} \qquad \forall\, \omega', \omega\, \in \Omega_g,\, \omega' \neq \omega, g \in \mathcal{G}_{t-1}, t \in \mathcal{T}\, t : \geq 2, j \in J \quad (b7a)$$

$$z^{\omega}_{j1} = z^{\omega'}_{j1} \qquad \forall\, \omega', \omega\, \in \Omega,\, \omega' \neq \omega, j \in J \quad (b7b)$$

$$z^{\omega}_{jt} \geq 0, \in \mathbb{Z}_+ \quad \forall t \in \mathcal{T},\, j \in J\, \omega \in \Omega \quad (b9)$$

Constraints $(b7a)$ and $(b7b)$ replace the original constraints in MA model $(s7)$ while $(b9)$ replaces $(s9)$. The rest of the model is unchanged.

#### 3.1.6.3. Expected Value Model, EV

The EV model is much like the basic deterministic model presented in section 3.1.1. However, the uncertain events are replaced by its expected value while the productivity and quality must determined by the the expected uncertainty. The expected uncertainty realization at moment $t$ can be obtained by equation 3.4. As $\xi^{\omega}_t$ could be $0$ or $1$, the expected rain event is the sum of the conditional probabilities of scenarios if the uncertainty realization is rain.

$$\bar{\xi}_t = \sum_{\omega \in \Omega'_t} w^{\omega}_t \xi^{\omega}_t = \sum_{\omega \in \Omega'_t} w^{\omega}_t \tag{3.4}$$

The expected productivity is calculated as:

$$\bar{\beta}_t = \hat{\beta} - \hat{\beta}(\bar{\xi}_t - \bar{\xi}_t \phi) \quad \forall\, t \in \mathcal{T} \tag{3.5}$$

Regarding quality, the expected value of quality is given in (3.6):

$$\bar{Q}_{jt} = (a_j t^2 + b_j t + c_j)(1 - u_j \sum_{t':1}^{t} \bar{\xi_{t'}}) \quad \forall\, j \in \mathcal{J}, t \in \mathcal{T} \tag{3.6}$$

#### 3.1.6.4. Wait-and-See model, WS

Finally, we present the Wait and See Model (WS) which is exactly the same as the MA, but all *nonanticipativity constraints* have been relaxed; it denotes the expected value of using the optimal solution for each scenario. So the WS model does not consider constraints $(s5),(s6)$ and $(s7)$. The rest of constraints and structure keeps similar. Under these conditions it can be expected that the following inequalities are satisfied $EV \leq MB \leq MA \leq WS$.

### 3.1.7. Model Comparison Metrics

To compare the solutions obtained by using different approaches, we will use three well known metrics: first, the *expected value of the solution of the EV* (EEV), second, *the expected value of perfect information* (EVPI) and third, *the value of the stochastic solution* (VSS) (Birge and Louveaux, 2011; Escudero et al., 2007). We will denote $\mathbb{U}_{MA}$ and $\mathbb{U}_{MB}$ as the solutions for the multistage stochastic models with and without recourse, respectevely. In the case of EV model we will apply the policy to the whole tree of scenarios which renders the EEV. The *Expected Value of Perfect Information* $EVPI$ is obtained by $EVPI = WS - \mathbb{U}_{MA}$ for the recourse model. We will also use the *Value of Stochastic Solution* $VSS$ formulated as $VSS = \mathbb{U}_{MA} - EEV$, and the *Value of Stochastic Solution without recourse* as $VSS_{MB} = \mathbb{U}_{MB} - EEV$. We use the Wait and See $WS$ model as an upper limit and report the results as a fraction of it.

Escudero et al. (2007) proposes a methodology for determining the VSS for multistage cases at each stage compared to our current approach of determining the difference between the expected values. We will compare them following the proposal by Upton (1994) which indicates a direct method, that is more intuitive for a manager.

To explore where value is generated in different approaches, we will look into four economical components of the problems: actual income as a percentage of the maximum income ($\mathbb{I}$), cost of labor as a percentage maximum income ($\mathbb{L}$), percentage deviation of grape quality from optimum conditions ($\mathbb{Q}$) and the percentage of unharvested grapes ($\mathbb{S}$).

$$\mathbb{I}_{\mathcal{M}} = \frac{\sum_{\omega \in \Omega} \sum_{t \in \mathcal{T}} \sum_{j \in \mathcal{J}} w^{\omega} Q^{\omega}_{jt} h^{\omega}_{jt} B_j}{\sum_{j \in \mathcal{J}} S_j B_j} \tag{3.7}$$

$$\mathbb{L}_{\mathcal{M}} = \frac{\sum_{\omega\in\Omega}\sum_{t\in\mathcal{T}}\sum_{j\in\mathcal{J}} w^{\omega}(C_E x_t + C_F y_t + C_P m_t)^{\omega}}{\sum_{j\in\mathcal{J}} S_j B_j} \tag{3.8}$$

$$\mathbb{Q}_{\mathcal{M}} = 1 - \frac{\sum_{\omega\in\Omega}\sum_{t\in\mathcal{T}}\sum_{j\in\mathcal{J}} w^{\omega} Q^{\omega}_{jt} h^{\omega}_{jt} B_j}{\sum_{\omega\in\Omega}\sum_{t\in\mathcal{T}}\sum_{j\in\mathcal{J}} w^{\omega} h^{\omega}_{jt} B_j} \tag{3.9}$$

$$\mathbb{S}_{\mathcal{M}} = 1 - \frac{\sum_{\omega\in\Omega}\sum_{t\in\mathcal{T}}\sum_{j\in\mathcal{J}} w^{\omega} h^{\omega}_{jt}}{\sum_{j\in\mathcal{J}} S_j} \tag{3.10}$$

The value of the percentage from maximum income ($\mathbb{I}_{\mathcal{M}}$) is constructed by dividing the expected income (considering grape quality) by the maximum possible income, given by the total amount of grapes multiplied by their price. For the value of cost of labor as a percentage of total income, ($\mathbb{L}_{\mathcal{M}}$), we determine the expected total cost, given by the cost of hiring, direct harvest cost, and the cost of worker termination, multiplied by the probability of each scenario divided by the total income. For the value of percentage deviation of grape quality from optimum conditions ($\mathbb{Q}_{\mathcal{M}}$) we do not include unharvested grapes, in order to capture just the quality effect. Finally, we indicate the final progress of harvesting by introducing the quotient of the expected final unharvested stock and the initial stock ($\mathbb{S}_{\mathcal{M}}$).

## 3.2. Results

In this section, we start by presenting the parameters and structure used in the computational experiments, and then present the main results of this work. In the results we will compare economic performances of the models for several conditions of rain and resource or labor ability. Secondly, we show the effect of quality in the contribution of each model. Immediately after, we discuss the way value is created, considering the harvest planning and real progress as central elements. To the end of this section, we introduce the effect of the rain probability transition matrix and the computational times.

### 3.2.1. Model Parameters

Table 3.1 presents the base parameters used in this work; most of them were obtained from the work by Ferrer et al. (2008).

Table 3.1. Model base parameters, most of them obtained from Ferrer et al. (2008)

| Model Parameter | Notation | | Value | Units |
|---|---|---|---|---|
| Grape Price | $B_j$ | Standard | 1.4 | \$/kg |
| | | High | 6 | \$/kg |
| Lay-off cost | $C_F$ | | 630 | \$/worker |
| Hiring cost | $C_E$ | | 420 | \$/worker |
| Maintain cost | $C_P$ | | 84 | \$/worker/period |
| Harvest cost | $C_H$ | | 0.28 | \$/harvested kg. |
| Initial harvest stock | $S_j$ | | 300,000 | kg. |
| Harvest Period | $n(\mathcal{T})$ | | 10 | days |
| Optimal Harvest periods | | | 5 and 6 | days |

To analyze the effect the parameters (e.g., workers ability, rain probability and transition, quality and the impact of rain) have on the models performance we will perform several optimizations with configurations or instances. The parameters used for each configuration or instance are summarized in Table 3.2. We tested a total of 2430 instances, so the total number of models optimized were 9720.

Table 3.2. Parameters for the different experiments

| Feature | Notation | Values | Units |
|---|---|---|---|
| Workers Ability | $\phi$ | 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 | – |
| Rain probability | $r_{01}$ | 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 | – |
| Transition factor | $\tau_{01}$ | 0.6, 0.8, 1.0, 1.2, 1.4 | – |
| Grape quality | Standard | a:-0.005, b:0.055, c:0.85 | \$ |
| | High | a:-0.04, b:0.44, c:-0.2 | \$ |
| Rain quality penalty | $u$ | 2%, 4%, 6% | % |

The models were implemented using Python , the model was written for PYOMO Python (Hart et al., 2017) and the optimization engine was GUROBI v. 8.1.0. The solution time was not limited, the optimality gap was set to 1% and the integrality parameter was the solver default. We used a laptop computer with an Intel Core processor i7-6700HQ CPU 2.60 GHz, with 32.0 GB of RAM memory running Windows 10.

### 3.2.2. Economic Performance

Table 3.3 shows the solutions obtained for different grape quality levels and rain probabilities ($r_{01}$) for the different models and the $EVPI$, $VSS_{MB}$ and $VSS$ metrics. In Table 3.4 we can observe the values in relative terms to the EEV model. Rain probabilities are defined as *low* when $r_{01} = 0.1$, *medium* when $r_{01} = 0.5$ and *high* when $r_{01} = 0.9$

Table 3.3. Solutions for WS, MA, MB and EEV models with $EVPI$, $VSS_{MB}$ and $VSS$ metrics, for three transition probabilities of rain, two grape qualities and low worker ability.

| Quality | $r_{01}$ | $WS$ | $\mathbb{U}_{MA}$ | $\mathbb{U}_{MB}$ | $EEV$ | $EVPI$ | $VSS$ | $VSS_{MB}$ |
|---|---|---|---|---|---|---|---|---|
| std | low | 241,458 | 237,383 | 237,286 | 233,970 | 4,075 | 3,413 | 3,315 |
| | medium | 191,532 | 182,634 | 181,925 | 175,544 | 8,898 | 7,090 | 6,380 |
| | high | 91,350 | 84,873 | 81,275 | 80,794 | 6,477 | 4,079 | 481 |
| hgh | low | 1,484,874 | 1,475,731 | 1,473,150 | 1,410,185 | 9,143 | 65,545 | 62,965 |
| | medium | 1,336,996 | 1,303,615 | 1,293,576 | 1,202,743 | 33,381 | 100,873 | 90,833 |
| | high | 1,113,867 | 1,082,632 | 988,975 | 966,236 | 31,235 | 116,396 | 22,739 |

Table 3.4. Solution values in relative terms to the EEV model for three transition probabilities of rain, two grape qualities and low worker ability.

| Quality | $r_{01}$ | $WS/EEV$ | $\mathbb{U}_{MA}/EEV$ | $\mathbb{U}_{MB}/EEV$ |
|---|---|---|---|---|
| std | low | 3.20% | 1.46% | 1.42% |
| | medium | 9.11% | 4.04% | 3.63% |
| | high | 13.07% | 5.05% | 0.60% |
| hgh | low | 5.30% | 4.65% | 4.46% |
| | medium | 11.16% | 8.39% | 7.55% |
| | high | 15.28% | 12.05% | 2.35% |

The MA model achieves its highest absolute value under low rain probability and high grape quality. However, the largest relative difference between the MA and EEV model is under high rain probability and high quality of grapes. This relative difference is followed by the high rain probability and standard quality of grapes scenario. If we observe the VSS for the MA model, we can see the same behavior; however in the case of high grape quality the largest VSS value is obtained under medium rain probability. In the case of the MB model, the largest relative difference between the model and EEV is under medium rain probability for both quality levels. A similar behavior is observed in the VSS, with the difference that highest value is obtained under medium rain probability and high quality of grapes.

Figure 3.5 presents the $VSS$ for two different grape qualities, for a range of rain probabilities and for three levels of resource ability. In the case of standard grape quality, we can observe a bell-curved shape for different rain probabilities, with a peak around 0.5 for high ability. If we observe the extreme probabilities of rain (0.1 and 0.9) the VSS value is smaller and the differences between the ability levels is also lower indicating that the use of MA approach produces less value than in the highly uncertain scenario (rain probability close to 0.5). In the case of high quality grapes, as the probability of rain increases, the benefit of using a MA approach is higher, indicating that for high quality grapes the use of an MA approach is more valuable when the probability of the negative event is higher, due to the value of the product. Hence quality and rain probability have a significant impact on $VSS$ while the ability of workers mitigates both effects.

In Figure 3.6 we can observe the $VSS_{MB}$. We can observe a similar behavior to the $VSS$ in the standard quality case, with a significant decrease in the value in the case of the high probability of rain event. In the case of the high quality grapes, we can observe a significant reduction in the $VSS_{MB}$ for the high probability of rain scenario, which indicates that the recourse action generates a significant portion of the of value in the case of unfavorable scenarios. For both cases the ability of the workers is able to mitigate the negative effect of the uncertainty and the quality of the grapes.

From these results, we can extract that the three factors which significantly affect the $VSS$ are: grape quality, labor ability, and rain probability.

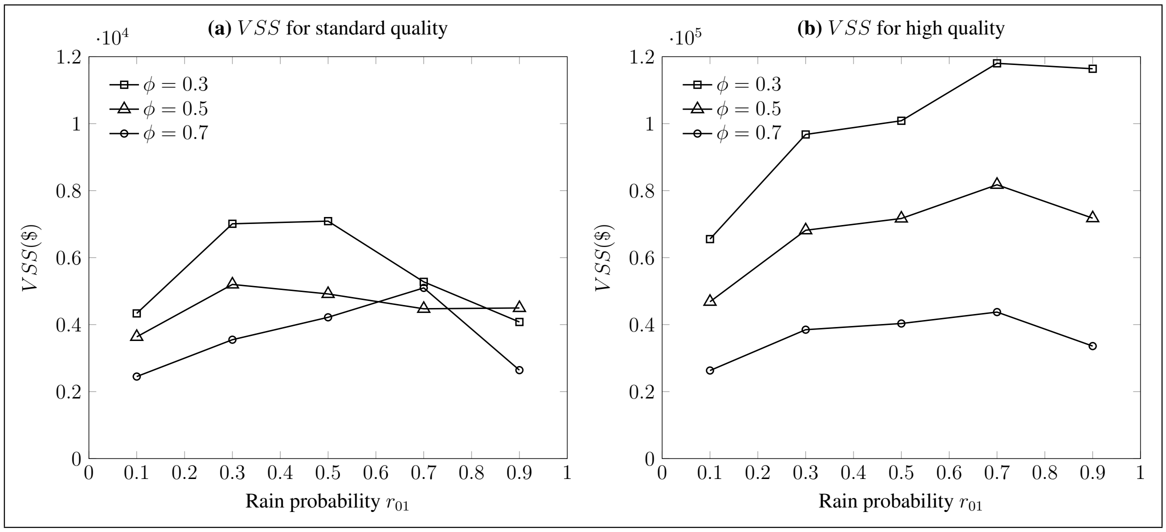



Figure 3.5. $VSS$ for two qualities of grapes, different rain transition probabilities and three levels of worker ability.

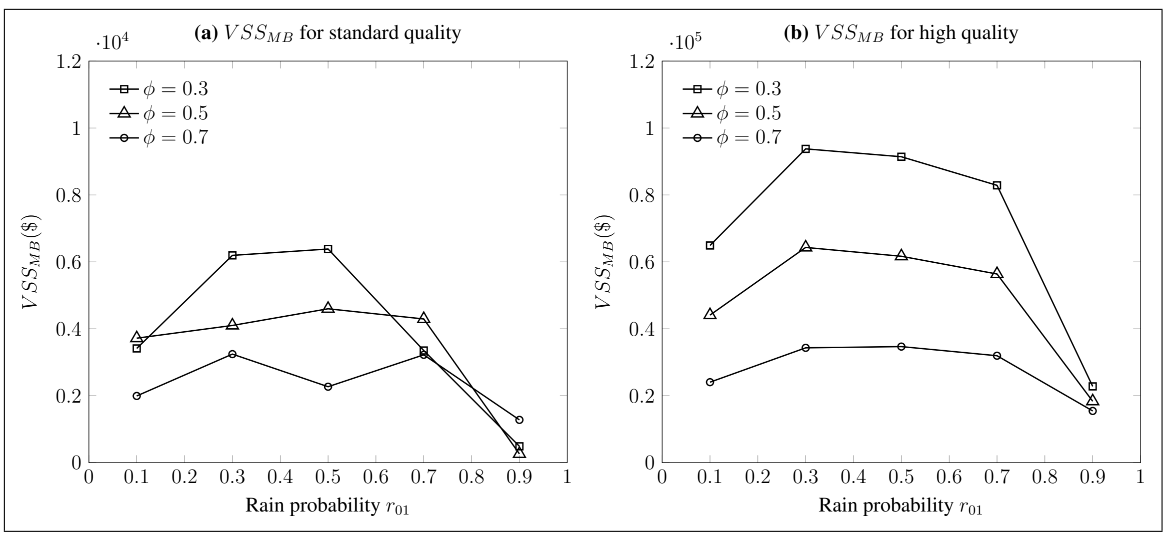



Figure 3.6. $VSS_{MB}$ for two qualities of grapes, different rain transition probabilities and three levels of worker ability.

### 3.2.3. Sources of Value

To determine the sources of value produced by the MA approach, we will present the results of four metrics: percentage deviation from maximum income ($\mathbb{I}$), cost of labor as a percentage of total income ($\mathbb{L}$), percentage of grape quality from optimum conditions ($\mathbb{Q}$), and the percentage amount of unharvested grapes ($\mathbb{S}$). Table 3.5 shows the results for the four metrics.

From the table, we can observe that when a rain event is highly uncertain (0.5), the maximum income percentage is higher for the MA than the EEV and MB. This indicates that the MA, under highly uncertain conditions, can capture more value from the grapes. This can be also reflected in the percentage of unharvested grapes, where the MA harvests nearly 100% while the EEV leaves a percentage behind. This difference is particularly clear for high quality grapes where the EEV loses over 3% of grapes across all levels of worker ability, while the MA loses less than 0.3%. The unharvested stock is similar in MA to MB. In the case of standard quality grapes, MA and EEV sacrifice almost the same percentage of quality. However, for high quality grapes, the MA sacrifices a higher percentage of quality than the EEV for all cases. This indicates that the MA model can capture more value by leaving fewer grapes behind. This is because the MA observes the posibility of rain in the future and reacts by advancing the harvest decisions of some blocks, while sacrificing the quality of the grapes, in order to leave fewer grapes behind. In the case of the MB, the impossibility to adjust the resource allocation, limits its reaction and so the value.

Table 3.5. Results for the four metrics for three levels of worker ability (flex), two levels of grape quality and two rain transition probabilities

| $r_{01}$ | Q | $\phi$ | I SR | I SNR | I EEV | L SR | L SNR | L EEV | $Q_L$ SR | $Q_L$ SNR | $Q_L$ EEV | S SR | S SNR | S EEV |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.5 | st | 0.3 | 89.8% | 89.4% | 85.2% | 46.3% | 46.1% | 43.4% | 7.8% | 7.8% | 7.8% | 2.44% | 2.8% | 7.0% |
| | | 0.5 | 91.5% | 91.1% | 87.5% | 43.5% | 43.1% | 40.7% | 8.0% | 7.9% | 8.4% | 0.49% | 1.1% | 4.1% |
| | | 0.7 | 91.4% | 91.1% | 88.9% | 39.9% | 40.2% | 38.5% | 8.4% | 8.1% | 8.8% | 0.21% | 0.7% | 2.3% |
| | hg | 0.3 | 89.8% | 89.5% | 82.1% | 17.4% | 17.6% | 15.3% | 10.1% | 10.2% | 7.7% | 0.10% | 0.3% | 10.2% |
| | | 0.5 | 90.9% | 90.2% | 85.5% | 15.3% | 15.2% | 14.0% | 9.0% | 9.7% | 8.2% | 0.09% | 0.2% | 6.3% |
| | | 0.7 | 91.1% | 90.9% | 87.9% | 13.8% | 14.0% | 13.0% | 8.8% | 9.0% | 8.7% | 0.13% | 0.2% | 3.3% |
| 0.9 | st | 0.3 | 81.3% | 78.9% | 81.3% | 61.0% | 59.5% | 62.1% | 11.9% | 11.2% | 11.9% | 6.89% | 9.9% | 6.8% |
| | | 0.5 | 87.5% | 85.5% | 84.0% | 51.0% | 50.0% | 48.6% | 12.4% | 11.8% | 12.6% | 0.08% | 2.7% | 3.4% |
| | | 0.7 | 86.8% | 87.0% | 86.6% | 42.6% | 43.1% | 43.0% | 12.7% | 12.0% | 11.7% | 0.46% | 0.9% | 1.6% |
| | hg | 0.3 | 82.1% | 75.5% | 74.4% | 22.0% | 20.5% | 20.7% | 17.7% | 20.0% | 14.4% | 0.15% | 4.5% | 11.2% |
| | | 0.5 | 87.0% | 83.9% | 81.6% | 18.9% | 18.8% | 17.5% | 13.0% | 15.2% | 12.3% | 0.00% | 0.9% | 6.0% |
| | | 0.7 | 86.6% | 86.9% | 84.2% | 14.9% | 16.1% | 14.4% | 13.3% | 12.9% | 13.0% | 0.12% | 0.2% | 2.8% |

When the rain probability is high (0.9), we can observe that the difference in the maximum income percentage is small between the MA and EEV. In the case of standard quality, those similarities are maintained across the metrics, indicating that the difference between the solutions given by the MA and EEV are similar. However, in the case of high quality grapes, the MA leaves fewer grapes behind while sacrificing quality. For both high uncertainty and high probability of rain, the MA model advances harvesting decisions in order to leave fewer grapes behind, while sacrificing quality. MB incomes are similar to the ones obtained by MA and EEV.

Worker ability has the effect of reducing the consequences of rain and the differences among models for all situations, allowing more value to be captured and costs to be reduced.

Quality and the decision to leave grapes behind play a significant role in the value generation process, especially in the moment the grapes are harvested. To corroborate that the MA approach foresees certain situations, and advances the harvest decision of some blocks, we will now compare the planning and execution of harvesting tasks to determine how the decision to harvest is handled by the MA, MB and EV approaches.

### 3.2.4. Planning and Execution of Harvesting

Tables 3.6 and 3.7 present the percentage of scenarios with harvest for each period which allow to observe the differences between the MA, MB and EV harvest decisions for two grape quality levels, two probabilities of rain (0.5 and 0.9) and three labor ability levels (Note that it is the planning step, so is EV and not EEV). The tables show the 10 harvest planning periods with their optimal harvest dates located in $t_5$ and $t_6$. The values inside represent, for each time period, the percentage of scenarios in which the model decided to harvest. In the case of the EV, since there is only one plan, the possible values are binary; however, the MA and MB models can make different harvest decisions according to previous conditions.

We can observe that for a standard quality level (Table 3.6), models tend to harvest during the entire planning period, thus reducing the number of workers required for harvest. As expected, the MA and MB tend to advance the harvest decision in comparison with the EV.

This is reflected in the final periods of harvest, where the EV performs harvest while the MA, in some scenarios, does not perform any harvest at all (values less than 1). When the probability of rain and its effect on harvest is more certain, increased to 0.9, the differences among approaches is reduced. Similarly, when labor ability is increased, the effect of rain and the differences between models are reduced.

In the case of high quality grapes (Table 3.7), we can now observe that the three models do not use the complete span of the planning periods, they concentrate near the optimal harvest time. When the probability of rain is highly uncertain (0.5), the EV approach concentrates the harvest very close to the optimal date in order not to be exposed to rain and hence lose quality. In the case of the MA model, since it accounts for the probability of no rain and can recourse in the event of rain, the decision to harvest is delayed to some post-optimal harvest periods in order to reduce the cost of harvesting. The MB model, delays even more the harvest decision since it does not have the possibility of a recourse action. When the probability of rain is increased to 0.9, we can observe that the number of scenarios in which the EV and MA continues with harvesting after the optimum window increases.

Table 3.6. Percentage of scenarios with harvest for each period for MA, MB and EV, three levels of worker ability, two rain transition probabilities and **standard quality** grapes.

| Q | $r_{01}$ | $\phi$ | M | $t_1$ | $t_2$ | $t_3$ | $t_4$ | $t_5$ | $t_6$ | $t_7$ | $t_8$ | $t_9$ | $t_{10}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| std | 0.5 | 0.3 | EV | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% |
| | | | SR | 100% | 100% | 100% | 100% | 100% | 100% | 98% | 93% | 76% | 59% |
| | | | SNR | 100% | 100% | 100% | 100% | 100% | 100% | 98% | 93% | 71% | 54% |
| | | 0.5 | EV | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% |
| | | | SR | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 95% | 72% | 51% |
| | | | SNR | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 95% | 72% | 47% |
| | | 0.7 | EV | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% |
| | | | SR | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 71% |
| | | | SNR | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 97% | 93% |
| | 0.9 | 0.3 | EV | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% |
| | | | SR | 100% | 100% | 100% | 100% | 100% | 100% | 99% | 97% | 81% | 81% |
| | | | SNR | 100% | 100% | 100% | 100% | 100% | 100% | 99% | 98% | 93% | 88% |
| | | 0.5 | EV | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% |
| | | | SR | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 97% | 81% | 74% |
| | | | SNR | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 99% | 94% | 84% |
| | | 0.7 | EV | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 0% |
| | | | SR | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 97% |
| | | | SNR | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 90% |

Table 3.7. Percentage of scenarios with harvest for each period for MA, MB, and EV, three levels of worker ability, two rain transition probabilities and **high quality** grapes.

| Q | $r_{01}$ | $\phi$ | M | $t_1$ | $t_2$ | $t_3$ | $t_4$ | $t_5$ | $t_6$ | $t_7$ | $t_8$ | $t_9$ | $t_{10}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| high | 0.5 | 0.3 | EV | 0% | 0% | 0% | 100% | 100% | 100% | 100% | 0% | 0% | 0% |
| | | | SR | 0% | 0% | 0% | 100% | 100% | 100% | 75% | 20% | 9% | 4% |
| | | | SNR | 0% | 0% | 0% | 100% | 100% | 100% | 62% | 31% | 16% | 8% |
| | | 0.5 | EV | 0% | 0% | 0% | 100% | 100% | 100% | 100% | 0% | 0% | 0% |
| | | | SR | 0% | 0% | 12% | 100% | 100% | 100% | 75% | 12% | 0% | 0% |
| | | | SNR | 0% | 0% | 0% | 100% | 100% | 100% | 75% | 37% | 19% | 10% |
| | | 0.7 | EV | 0% | 0% | 0% | 100% | 100% | 100% | 100% | 100% | 0% | 0% |
| | | | SR | 0% | 0% | 0% | 100% | 100% | 100% | 100% | 7% | 1% | 0% |
| | | | SNR | 0% | 0% | 0% | 100% | 100% | 100% | 87% | 44% | 22% | 13% |
| | 0.9 | 0.3 | EV | 0% | 0% | 100% | 100% | 100% | 100% | 100% | 0% | 0% | 0% |
| | | | SR | 0% | 0% | 100% | 100% | 100% | 97% | 73% | 66% | 0% | 0% |
| | | | SNR | 0% | 0% | 100% | 100% | 99% | 98% | 87% | 79% | 71% | 69% |
| | | 0.5 | EV | 0% | 0% | 0% | 100% | 100% | 100% | 100% | 100% | 0% | 0% |
| | | | SR | 0% | 0% | 0% | 100% | 100% | 99% | 73% | 0% | 0% | 0% |
| | | | SNR | 0% | 0% | 0% | 100% | 100% | 99% | 89% | 80% | 72% | 65% |
| | | 0.7 | EV | 0% | 0% | 0% | 100% | 100% | 100% | 100% | 100% | 0% | 0% |
| | | | SR | 0% | 0% | 0% | 100% | 100% | 100% | 100% | 48% | 0% | 0% |
| | | | SNR | 0% | 0% | 0% | 100% | 100% | 100% | 90% | 81% | 73% | 65% |

Table 3.8 shows the average percentage of accumulated harvest for standard quality grapes. We can observe that both approaches use almost the complete time span of harvest, with an average of 10% of harvest for each day. In the case of EEV, the solution uses a stable harvest approach of 10% to reduce the harvest cost, except in a scenario with a high probability of rain and high labor ability, in which case the harvest is increased to 11.1% to shorten the harvest period. As labor skill increases the percentage of unharvested grapes is significantly reduced for EEV and MB models, while in the case of the MA model the reduction is more stable as the ability is increased. For high probability of rain, the MA and MB advances the harvest decision so that around 7% more grapes are harvested by the end of the optimal period ($t_5$ and $t_6$). This is done to avoid damage caused by rain in the final periods and, hence, it allows the model to leave fewer grapes unharvested.

If we look at high quality grapes (Table 3.9), when the probability of rain is highly uncertain (0.5), we can observe that the EEV leaves on average a significant volume of grapes unharvested, compared to the MA. This is because the EEV approach concentrates the harvest very close to the optimal dates, in a small time window of 4 to 5 days, in order to reduce the quality degradation of the grapes. Unfortunately, when the harvest plan is implemented in the EEV, there are some scenarios of rain in which the workers cannot harvest the amounts planned and, due to its inflexible nature, all unharvested grapes are left behind. Since the MA approach adjusts the harvest plan and labor to the prevailing rain conditions, the harvest is extended to the end of the period, which leaves fewer grapes unharvested, although their quality is reduced. Thus, the ability of the MA model to account for different rain scenarios allows the harvest to be delayed and, in the event of rain, adjusts the labor requirement in order to leave fewer grapes unharvested, compared to the EEV. When the event of rain is more likely (0.9) we can observe a similar behavior as in the highly uncertain case, though with a smaller gap in the amount of unharvested grapes between the EV and MA approaches. Furthermore, as expected, the ability of the workers tends to reduce the differences between the models. Due to the recourse actions, the MA approach advances on average more harvest in the optimum days compared to the MB and EEV, allowing to capture both quality and volume.

From these results, we can see that the probability of rain also has an important effect on the approaches decisions, and the generation of value. In the following section we will explore the effect of rain probabilities.

Table 3.8. Average percentage volume of accumulated harvest for each period for EEV, MB and MA, for three levels of worker ability, two rain transition probabilities and standard quality grapes.Final column indicates the average percentage volume of grapes not harvested.

| Q | $r_{01}$ | $\phi$ | M | $t_1$ | $t_2$ | $t_3$ | $t_4$ | $t_5$ | $t_6$ | $t_7$ | $t_8$ | $t_9$ | $t_{10}$ | Left |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| std | 0.5 | 0.3 | EEV | 10.1% | 20.2% | 30.3% | 40.4% | 50.5% | 60.6% | 70.6% | 80.0% | 87.5% | 93.3% | 6.7% |
| | | | SR | 11.3% | 22.6% | 33.9% | 45.3% | 56.8% | 68.2% | 79.0% | 87.9% | 93.8% | 97.3% | 2.7% |
| | | | SNR | 11.3% | 22.6% | 34.0% | 45.5% | 57.0% | 68.5% | 79.3% | 88.1% | 93.9% | 97.5% | 2.5% |
| | | 0.5 | EEV | 10.0% | 20.0% | 30.0% | 40.0% | 50.0% | 60.0% | 70.0% | 80.0% | 89.3% | 95.9% | 4.1% |
| | | | SR | 11.4% | 23.0% | 34.5% | 46.0% | 57.5% | 69.0% | 80.5% | 90.4% | 96.4% | 98.9% | 1.1% |
| | | | SNR | 11.2% | 22.7% | 34.3% | 45.9% | 57.5% | 69.1% | 80.6% | 90.6% | 96.7% | 99.6% | 0.4% |
| | | 0.7 | EEV | 10.0% | 20.0% | 30.0% | 40.0% | 50.0% | 60.0% | 70.0% | 80.0% | 89.9% | 97.9% | 2.1% |
| | | | SR | 11.1% | 22.4% | 33.5% | 44.6% | 55.7% | 66.8% | 77.9% | 88.9% | 96.9% | 99.3% | 0.7% |
| | | | SNR | 10.9% | 21.8% | 32.7% | 43.6% | 54.5% | 65.4% | 76.3% | 87.2% | 96.3% | 99.7% | 0.3% |
| | 0.9 | 0.3 | EEV | 10.0% | 20.0% | 30.0% | 40.0% | 50.0% | 59.9% | 69.6% | 78.7% | 86.4% | 93.4% | 6.6% |
| | | | SR | 10.0% | 20.0% | 30.0% | 40.0% | 49.9% | 59.7% | 69.1% | 77.6% | 84.6% | 90.3% | 9.7% |
| | | | SNR | 10.0% | 20.0% | 30.0% | 40.0% | 49.9% | 59.8% | 69.5% | 78.6% | 86.3% | 93.3% | 6.7% |
| | | 0.5 | EEV | 10.0% | 20.0% | 30.0% | 40.0% | 50.0% | 60.0% | 70.0% | 79.9% | 89.4% | 96.8% | 3.2% |
| | | | SR | 11.0% | 22.0% | 33.0% | 44.0% | 55.0% | 66.0% | 76.9% | 87.0% | 94.2% | 97.4% | 2.6% |
| | | | SNR | 11.0% | 22.0% | 33.0% | 44.0% | 54.9% | 65.9% | 76.9% | 87.5% | 96.1% | 100.0% | 0.0% |
| | | 0.7 | EEV | 11.1% | 22.2% | 33.3% | 44.4% | 55.5% | 66.6% | 77.7% | 88.8% | 98.4% | 98.4% | 1.6% |
| | | | SR | 10.9% | 22.0% | 33.1% | 44.2% | 55.3% | 66.4% | 77.5% | 88.5% | 96.9% | 99.2% | 0.8% |
| | | | SNR | 10.5% | 21.0% | 31.5% | 42.0% | 52.5% | 63.0% | 73.5% | 84.0% | 94.3% | 99.5% | 0.5% |

Table 3.9. Average percentage volume of accumulated harvest for each period for EEV, MB, and MA, for three levels of worker ability, two rain transition probabilities and high quality grapes. Final column indicates the average percentage volume of grapes not harvested.

| Q | $r_{01}$ | $\phi$ | M | $t_1$ | $t_2$ | $t_3$ | $t_4$ | $t_5$ | $t_6$ | $t_7$ | $t_8$ | $t_9$ | $t_{10}$ | Left |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| high | 0.5 | 0.3 | EEV | 0.0% | 0.0% | 0.0% | 25.0% | 50.0% | 73.1% | 89.9% | 89.9% | 89.9% | 89.9% | 10.1% |
| | | | SR | 0.0% | 0.0% | 0.0% | 28.3% | 56.6% | 82.1% | 93.8% | 97.9% | 99.3% | 99.8% | 0.2% |
| | | | SNR | 0.0% | 0.0% | 0.0% | 24.9% | 51.8% | 78.8% | 94.4% | 98.6% | 99.8% | 100.0% | 0.0% |
| | | 0.5 | EEV | 0.0% | 0.0% | 0.0% | 25.0% | 50.0% | 75.0% | 93.8% | 93.8% | 93.8% | 93.8% | 6.2% |
| | | | SR | 0.0% | 0.0% | 0.0% | 25.0% | 53.1% | 81.2% | 95.3% | 98.8% | 99.7% | 99.9% | 0.1% |
| | | | SNR | 0.0% | 0.0% | 0.0% | 25.0% | 53.1% | 81.2% | 97.6% | 99.9% | 99.9% | 99.9% | 0.1% |
| | | 0.7 | EEV | 0.0% | 0.0% | 0.0% | 24.9% | 49.8% | 74.7% | 96.4% | 96.6% | 96.6% | 96.6% | 3.4% |
| | | | SR | 0.0% | 0.0% | 0.0% | 28.3% | 56.6% | 84.9% | 97.6% | 99.5% | 99.8% | 99.9% | 0.1% |
| | | | SNR | 0.0% | 0.0% | 0.0% | 27.3% | 54.7% | 82.1% | 99.2% | 99.8% | 99.8% | 99.8% | 0.2% |
| | 0.9 | 0.3 | EEV | 0.0% | 0.0% | 20.0% | 39.9% | 59.3% | 76.9% | 88.7% | 88.7% | 88.7% | 88.7% | 11.3% |
| | | | SR | 0.0% | 0.0% | 18.3% | 37.8% | 55.9% | 72.1% | 82.3% | 88.8% | 92.8% | 95.4% | 4.6% |
| | | | SNR | 0.0% | 0.0% | 16.1% | 37.0% | 57.6% | 75.9% | 89.4% | 99.9% | 99.9% | 99.9% | 0.1% |
| | | 0.5 | EEV | 0.0% | 0.0% | 0.0% | 24.9% | 49.8% | 74.3% | 93.8% | 93.9% | 93.9% | 93.9% | 6.1% |
| | | | SR | 0.0% | 0.0% | 0.0% | 27.3% | 54.8% | 79.6% | 90.8% | 95.8% | 98.0% | 99.0% | 1.0% |
| | | | SNR | 0.0% | 0.0% | 0.0% | 27.5% | 55.0% | 81.7% | 99.9% | 99.9% | 99.9% | 99.9% | 0.1% |
| | | 0.7 | EEV | 0.0% | 0.0% | 0.0% | 24.9% | 49.8% | 74.7% | 97.0% | 97.1% | 97.1% | 97.1% | 2.9% |
| | | | SR | 0.0% | 0.0% | 0.0% | 30.4% | 60.8% | 89.3% | 97.0% | 99.1% | 99.6% | 99.7% | 0.3% |
| | | | SNR | 0.0% | 0.0% | 0.0% | 26.0% | 51.8% | 77.9% | 99.7% | 99.8% | 99.8% | 99.8% | 0.2% |

### 3.2.5. Rain Probability Effects

Since we modeled the probability of rain using a Markov process, the rain probability effect can be divided into two components: first, the transition probability of changing from a non-rainy day into a rainy day, which we defined as $r_{01}$ and, second, the probability of rain to continue from one day to the next, which we stated as $r_{11}$. Figures 3.7 and 3.8 present the percentage value of $VSS$ over the $EEV$ for low and high labor ability, for a given set of $r_{01}$ and percentage changes in the transition factor $\tau_{01}$, being $0\%$ when $r_{01} = r_{11}$.

Figure 3.7 $(a)$ shows that for standard quality and low labor ability, the $VSS$ value against the $EEV$ is not significantly affected by the percentage changes in the quotient $r_{11}/r_{01}$ for medium and low rain probabilities ($r_{01}$). However, when rain probability is high (0.9) and as the quotient $r_{11}/r_{01}$ increases, the value of using an stochastic approach is increased. Figure 3.8 $(a)$, presents the results for standard quality and high ability, in all cases the value of using the stochastic approach is reduced if we compare with the low ability. When the probability of rain is low ($r_{01} = 0.1$), the $VSS$ is unaffected by changes in the quotient $r_{11}/r_{01}$. When rain probability is very high, $r_{01} = 0.9$ and high $r_{11}/r_{01}$ quotient, the $VSS$ has no difference with the $EEV$. When $r_{01} = 0.5$ and as the quotient $r_{11}/r_{01}$ increases, the value of using an stochastic approach is increased. In 3.7 $(b)$ and 3.8 $(b)$, we can observe the same behavior, but the differences in the $VSS$ are increased due to the quality of the grapes, more in the case of low ability than in the high ability.

### 3.2.6. Value of Worker Reassignment Flexibility

As it was mentioned at the beginning of the work, the non-recourse (MB) and recourse (MA) modeling approaches of grape harvest differ, since the latter can reassign labor once the state of the nature reveals. The possibility or flexibility to reassign after the uncertainty reveals itself gives a performance advantage to the MA approach. We will now look at the value differences between both models, under various scenarios, in order to determine under which conditions the recourse action renders more value. In Figures 3.9 and 3.10 we can observe the percentage value difference between MA and MB for low and high

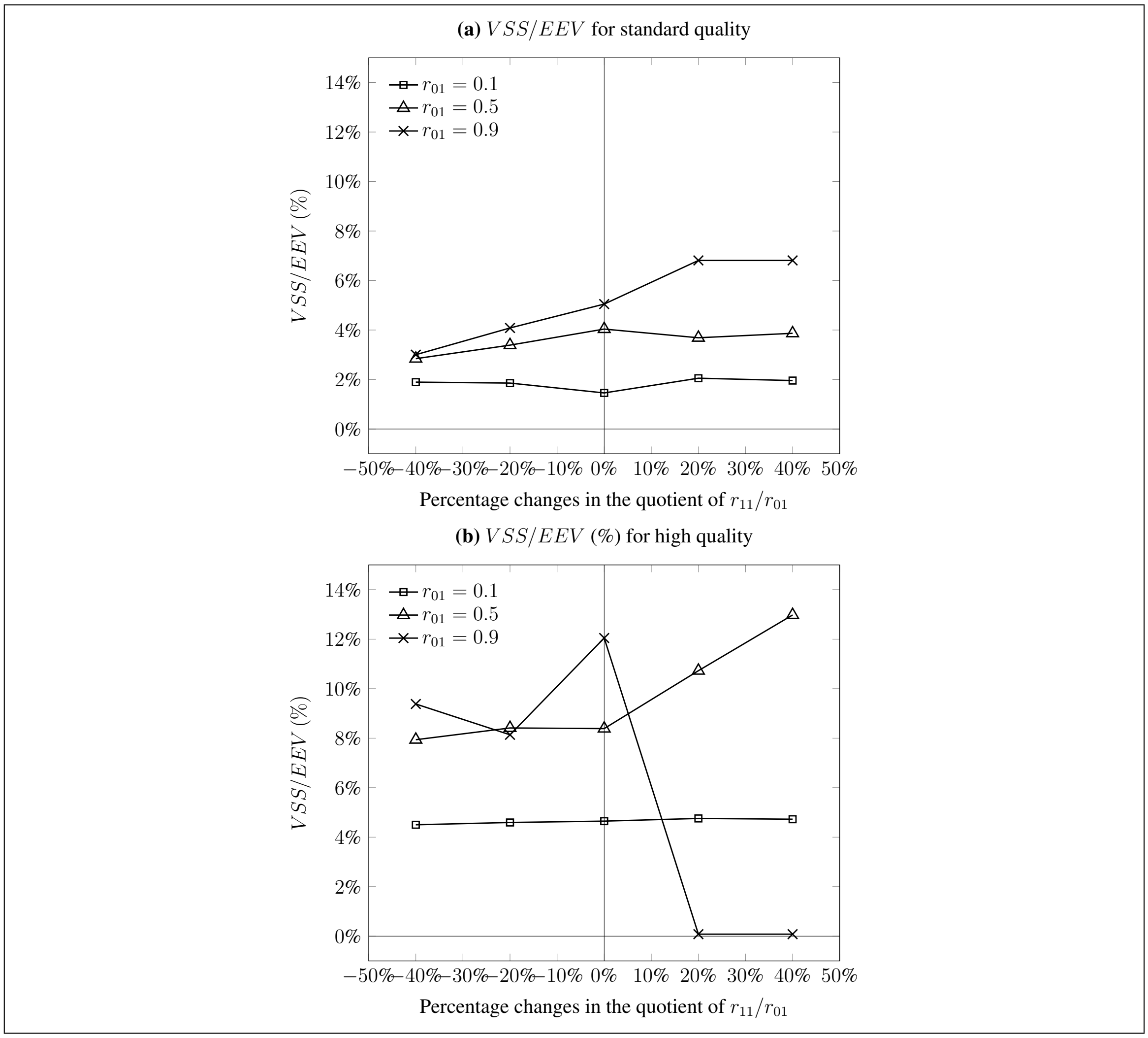



Figure 3.7. $VSS$ percentage for **low labor ability**, different probabilities of rain ($r_{01}$) and percentage changes in the quotient $r_{11}/r_{01}$

labor ability, respectively, for a given set of $r_{01}$, and percentage changes in the quotient $r_{11}/r_{01}$.

For low labor ability and both standard and high quality, we can observe that the value generated by the MA approach is significant when the probability of rain is high ($r_{01} = 0.9$) and the gap is enlarged as the quotient $r_{11}/r_{01}$ increases. When the rain event is improbable $r_{01} = 0.1$ or uncertain $r_{01} = 0.5$ the differences between both approaches is not significant for both cases. The gap increases according $\tau_{01}$ is bigger. In Figure 3.10 we can observe that when the resource ability is high the effects of the recourse actions

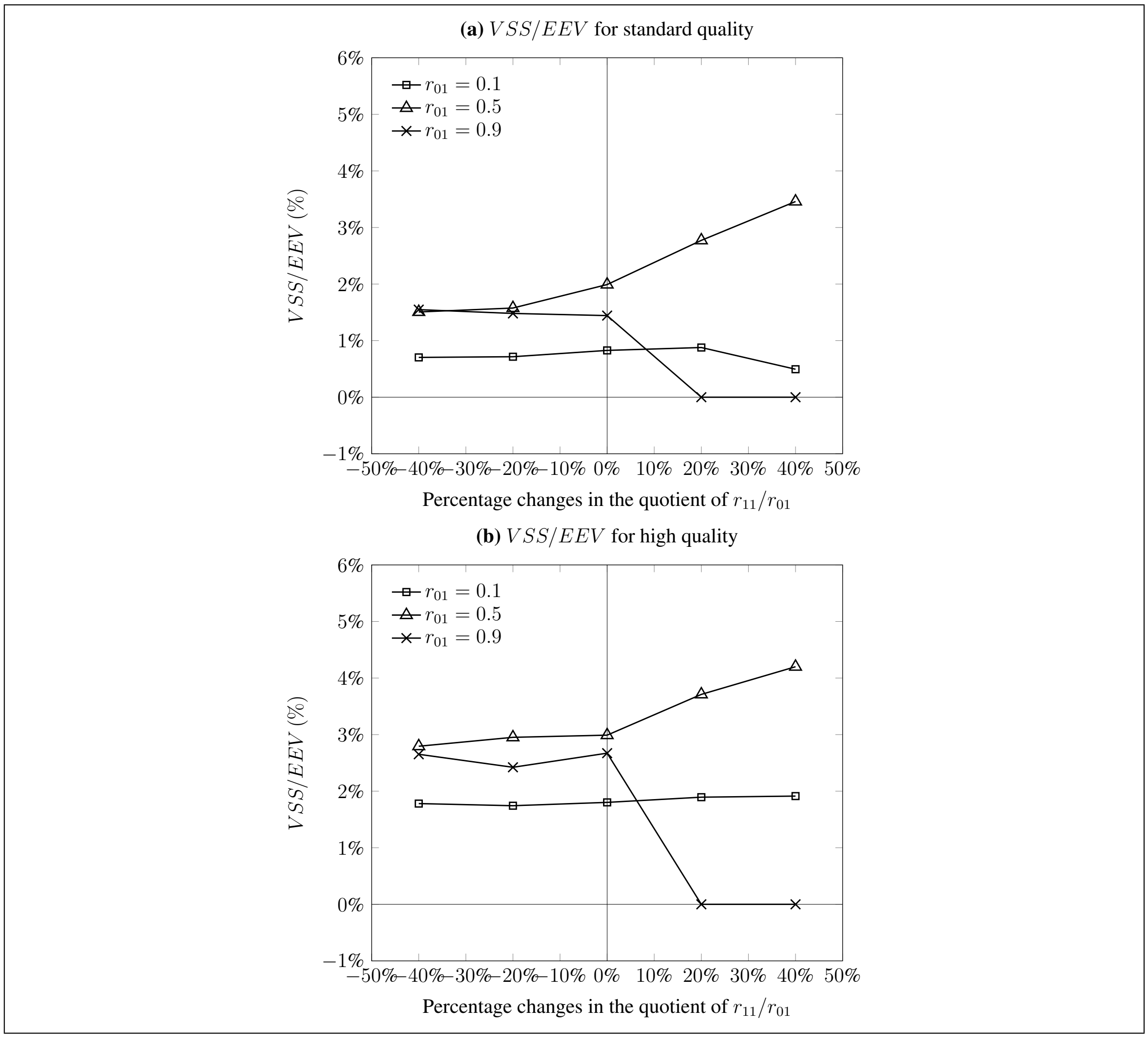



Figure 3.8. $VSS$ percentage for **high labor ability**, different probabilities of rain ($r_{01}$) and percentage changes in the quotient $r_{11}/r_{01}$

are diluted, however under high rain probability we can observe that the benefits of using the recourse approach are increased.

### 3.2.7. Computational Times and the Effect of Instance Size

In the previous sections we have seen that under certain conditions the MA generates better solutions than the EV approach. This is because, unlike the EV, it accounts for the uncertainty in weather conditions and does not prescribe a single solution but a set, or tree, of solutions which are conditional to how the states of nature are revealed. Obtaining a set, or tree, of solutions does not come without cost; simultaneous optimization models need

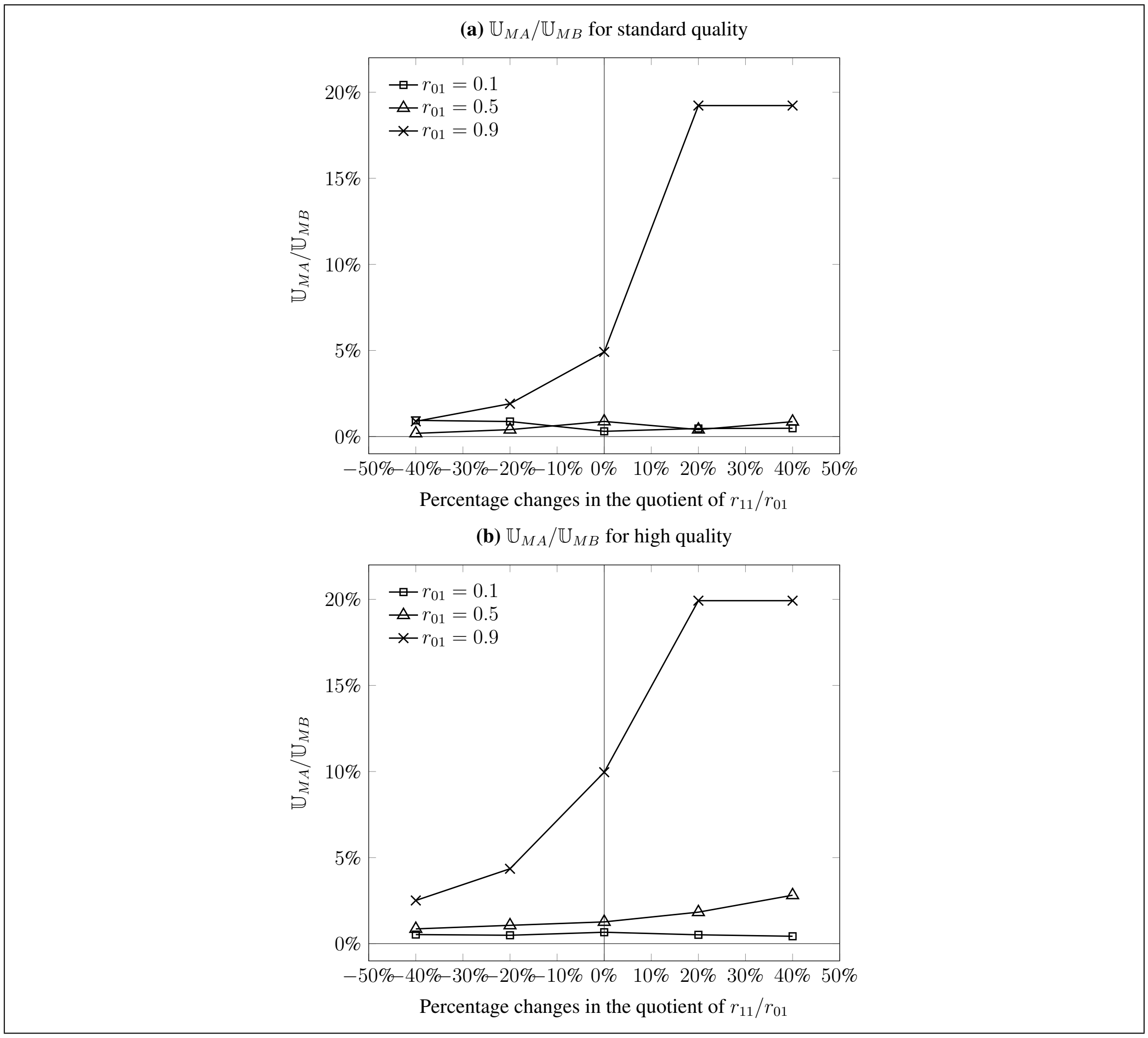



Figure 3.9. Percentage differences MA and MB for a **low labor ability**, different probabilities of rain ($r_{01}$) and percentage changes in the quotient of $r_{11}/r_{01}$.

to be solved for each scenario and must also account for *non-anticipativity principle*. This puts a significant computational burden on the MA approach, which grows exponentially with the number of periods.

To determine the computational effect that the number of periods has on the MA approach against the EV, we run a number of experiments in which we increase the number of periods and record the difference in time required to obtain an optimal solution for the MA and EV. In Figure 3.11 we can observe the results of these experiments. The computation time is shown in seconds and is the difference between MA and EV solving time; the

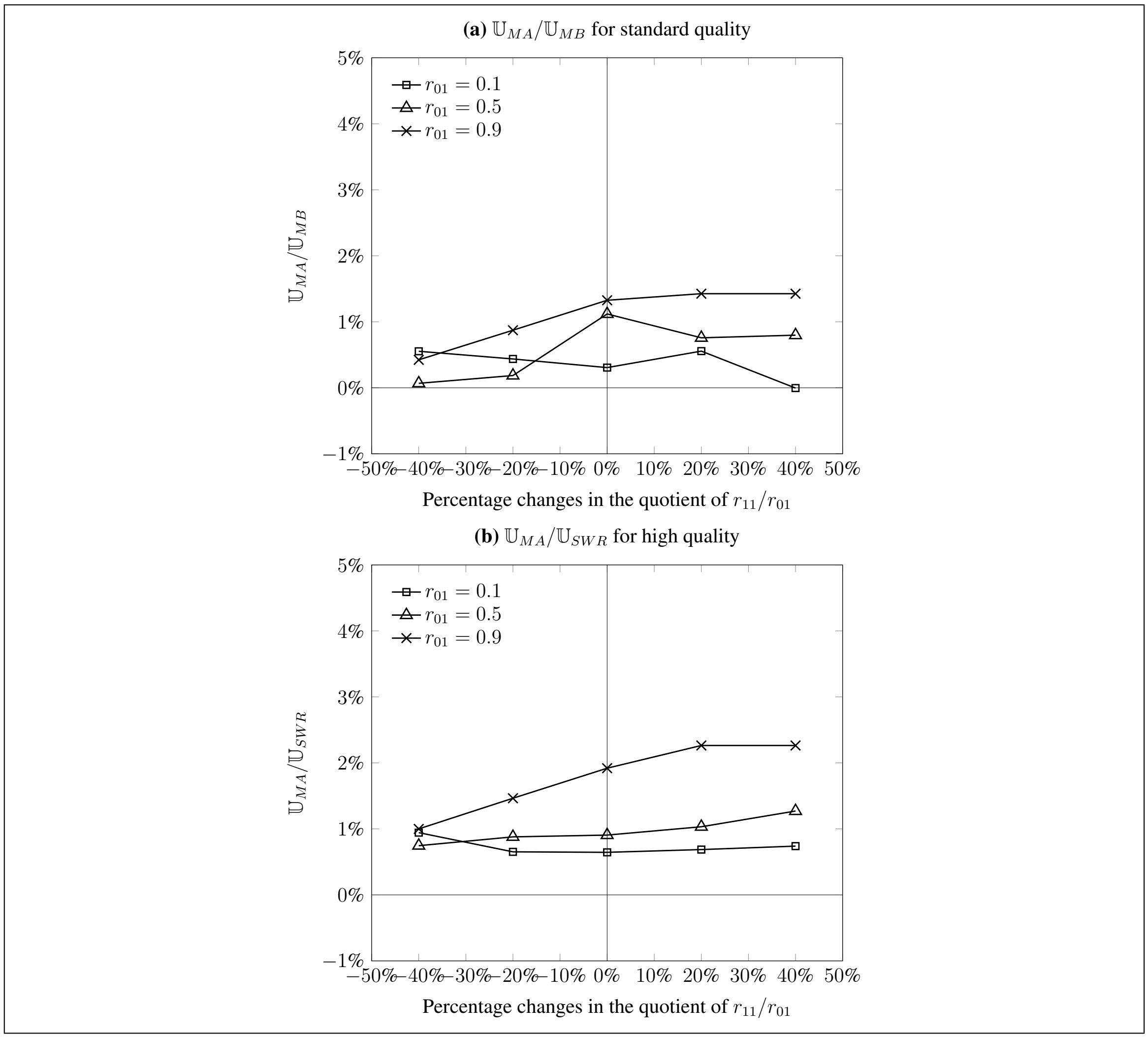



Figure 3.10. Percentage differences between flexible MA and MB for **high worker ability**, different probabilities of rain ($r_{01}$) and percentage changes in the quotient of $r_{11}/r_{01}$.

result is similar to previous literature where, as the quantity of instances increases, the computation time grows exponentially for the MA compared to the EV.

## 3.3. Discussion and Conclusions

We presented two multistage stochastic optimization models for planning wine harvest operations that accounts for rain uncertainty and quality degradation of the product. The first one, the MA model with recourse actions which can reassign labor once the state of the nature reveals; the second, the MB model which cannot reassign labor once the state

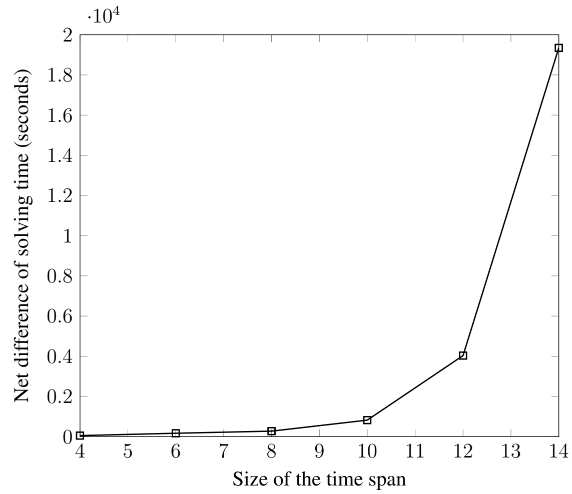



Figure 3.11. Differences in net solving time for different sizes (periods of harvesting) between the MA and EV approaches.

of the nature reveals. Both models were developed to analyze and determine the value of using a stochastic approach against an expected one and the value of the recourse actions, under different scenarios. The solutions obtained using the MA model were significantly different to the one obtained using an expected value approach and provides better results when tested against all possible scenarios. This result is similar to what Ahumada et al. (2012) concluded when applying a two-stage stochastic optimization approach.

The MA model can adjust the labor requirements and assignment as more information is available, and its decision process allows flexibility in whether to advance or defer harvesting. Results indicate that we can produce solutions that would generate up to 8% more value by using a MA approach rather than a EV. In some specific cases, MB produces a solution which is 12% lower in value than its EV counterpart, specially as effect of the structure and the nonexistence of recourse actions. The parameters that affect the harvest plan are: grape quality, rain probability, worker ability and decision flexibility.

As grape quality improves, the value of using a MA approach increases for every instance. The source of the value produced by using the MA comes from the ability of the model to not prescribe a single solution, but a set, or tree, of solutions which are conditional to the state of nature. The value produced by implementing the MA solution comes from two main actions the models performs: first, the model leaves fewer grapes unharvested while maintaining a similar harvest cost as the EV and, second, it does not sacrifice the quality of the grapes. These two sources of values are reflected in the two harvest actions the model takes: first, it advances harvest decisions when the quality is not significantly affected by moving from the optimal harvest date (standard quality grapes), especially when the probability of rain is uncertain (0.5). Second, when the quality of the grapes is high, the model harvests as many grapes as possible within the optimal window, while also postponing a small part of the harvest in the event that it does not rain.

We modeled the rain effect using a Markov process composed of two probabilities: first, the transition probability of changing from a non-rainy day into a rainy day and second, the probability of rain to continue from one day to the next. When rain probability is uncertain and as the transition probability is closer to 0.5, using a MA approach significantly renders more value than the EEV approach. This difference is reduced as the transition probability is near the extremes (0.1 and 0.9).

When both the transition probability of rain (0.9) and the recurrent probability of rain are high, the MA approach renders similar value than the EEV. If we compare the MA with MB approach, the latter produces less value due to the inability of MB model to adjust worker assignment according to changes in the state of nature. This result is similar to what Goyal and Netessine (2007) found in manufacturing service, where they conclude that flexibility should be favored in highly uncertain environments.

Finally, we integrated worker ability, which reflects the effect of rain on productivity, and decision making flexibility into the model. As worker ability increases, the effect of rain on the harvest is reduced, as is the value of using the MA and MB approach. The use of an stochastic approach adds complexity into the process, since the formulation, data collection and implementation are more elaborate. Furthermore, the computational time grows exponentially with the number of instances, which can be a deterrent from using such an approach.

The main managerial insight obtained from this research is identifying under which conditions the use of the MA approach renders more value than the EV single solution approach. In Table 3.10 we can observe how the $VSS$ for the MA and MB models compare to the EEV. The MA model always generates more value than the MB approach. Specially, if

Table 3.10. Values for $VSS/EEV$ and $VSS_{MB}/EEV$ for different scenarios

| | | SR | | | | SNR | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Standard Quality | | High Quality | | Standard Quality | | High Quality | |
| $r_{11}/r_{01}$ | $r_{01}$ | Low Ability | High Ability | Low Ability | High Ability | Low Ability | High Ability | Low Ability | High Ability |
| < 1 | 0.1 | 2% | 0% | 2% | 2% | 2% | 0% | 2% | 2% |
| | 0.5 | 4% | 2% | 8% | 2% | 4% | 2% | 6% | 2% |
| | 0.9 | 4% | 0% | 8% | 2% | 2% | 2% | 6% | 2% |
| 1 | 0.1 | 2% | 2% | 6% | 2% | 2% | 0% | 4% | 2% |
| | 0.5 | 2% | 2% | 7% | 2% | 2% | 2% | 6% | 2% |
| | 0.9 | 6% | 0% | 8% | 2% | 0% | 2% | 2% | 2% |
| > 1 | 0.1 | 2% | 2% | 4% | 2% | 2% | 0% | 4% | 2% |
| | 0.5 | 4% | 2% | 8% | 4% | 4% | 2% | 6% | 4% |
| | 0.9 | 6% | 0% | 0% | 0% | -6% | 0% | -6% | 0% |

we are dealing with grapes of high quality and low ability, the benefits of using the $MA$ are significant and positive for all parameters, therefore the use of this approach is recommended. Moreover, the $VSS_{MB}$ presents very poor performance when the rain probability is high ($r_{01}$) and the quotient $r_{11}/r_{01} > 1$, results can even be lower than the EEV. If ability is high, the behavior is quite similar, so using MA may not be critical, and MB could be a good option. When Quality is standard, the price of the grape limits the value of the whole system, so the benefits of using an stochastic approach are not as significant as the high quality case. For low abilities, MA renders better than MB compared to the case of high ability in which the benefits of using an stochastic approach are not so significant. Hence under high quality grapes, low ability of labor and high probabilities of rain, the decision maker should definitely use an stochastic approach. The same would hold for standard quality of grapes, however the economic benefits would be smaller. Under standard quality and high ability of labor, the benefits are significantly reduced and so the decision maker can rely on a standard expected value approaches

Our stochastic modeling approach has several limitations. First, to model the event of rain we use a Markov chain binomial approach, which does not account for the intensity of the event. The intensity may be useful to indicate intermediate effects on the fruit or soils conditions, important for machine jobs. This should be addressed by either increasing the possible states or by generating an intensity feature for the rain event. Second, we only use one type of harvest resource (labor); in reality, managers can also use machinery to perform the harvest, or different expertise teams. Third, we did not set an external constraint that accounts for how the manager needs to consider the daily capacity of the cellar to receive grapes or, in the case of a third-party buyer, the maximum level that can be sold. Fourth, in modeling the effect of quality degradation we account for the effect of previous rain and use a symmetrical concave function, while in the work by Arnaout and Maatouk (2010) they use a non-symmetrical quality degradation function. Finally, since our MS model is based on a Mixed-Integer Programming approach with a binomial scenario generation scheme, the computational times grow exponentially with the number of instances, making the model solvable only for harvesting problems with fewer instances.

# Chapter 4

# The value of decisions and labor ability as flexibility sources

In the previous chapter, we compared a simplified model, EV, with a stochastic approach, MS, applied in the harvesting of grape wine. The aim was to understand the contribution of complex models when flexible resources are present, in context of uncertainty. The uncertainty refers to rain conditions, being a binary variable, and its impact on quality and productivity. Quality is deteriorated if rain happens; it limits the quality of the block, so the final income decreases. About productivity, harvest resource is represented by manpower, in a context of manual picking. The resource has a nominal productivity that is the maximum that can be reached; however, depending on the flexibility for working under rain conditions, the productivity in the rainy periods will be affected. Highly flexible resources suffer slight penalization in rainy conditions. Three different ripening patterns were examined under varying sequences of the optimum maturity days. We concluded that very skilled resources gave stability to the decisions and the EV model has an acceptable performance compared to MS. If the uncertainty increases, the gap between EV and MS is small only for highly flexible resources. When the ability decays, the advantage for MS increases considerably. Finally, when rain is highly probable, EV behaves better than the MS without recourse action; but if recourse action is allowed, then MS recover the advantage.

In general terms, the ability to harvest under the rain is a kind of flexibility. It gives the capacity to react to the negative event with a correct answer in some degree. Nominal productivity is also flexibility, because it gets better the performance base. As was indicated by Buzacott and Mandelbaum (2008), the number of decisions and their stages are ways of flexibility, even when they are not always considered like this in operation research. MS formulation is more flexible than EV approach because it has detailed information to make decisions. But the created value does not justify always the complexity of MS. This problem is part of a bigger one: flexibility value is complex to understand and predict (Rogalski, 2011).

In this chapter we will extend the Chapter 3 to other ways of flexibility. The first source of flexibility to value is the allocation epoch. In Chapter 3 the allocation is cost-free, and we tested two options limited by the information about the uncertainty realization. If any of decisions were made before uncertainty is revealed, the MS model is called MB. If any of decisions are made after the uncertainty is revealed, the MS variant is called MA. We observe that the last option creates more value; in the free-charge model, the extra value of MS model, keeping all the rest of conditions céteris páribus, is the willingness to pay for the flexibility of delaying the decision. Delaying decisions is linked to the information availability, so, it is a way of value the information update. Now, we propose to compare three variants of MS models, MB, MA, and after-and/or-before (MC). Making decision before is cheaper than after, because the risk of making before is compensated with minor costs (the compensation is not exactly). The after-and/or-before alternative offers the opportunity of a later correction if decision was made before, or just to make the decision after uncertainty realization. The worst situation about cost is to decide before and change the decision then, assuming both costs. To refer to exact instant of the decisions, we observe that the limit is the uncertainty realization; henceforth, we will say that MB and MA make at least one of the decisions in different *epochs* in the same periods.

The other source of flexibility is the harvest resource. We explore here the value creation considering first, a single resource with variable ability to harvest under rain, and second, providing to the problem of two different resources, experts and rookies. The ability to harvest under the rain is similar but experts are more productive. This extra capacity has a cost. Choosing the team components is a flexibility for the farmer; intuitively expert should be better, but their cost is higher so the obvious decision are not such in uncertain contexts. In order to understand what the cost impact is on team selection under uncertainty, we carried out a sensitivity analysis of cost and productivity.

To monitor the performance gaps, we follow the metrics VSS and EVPI, previously defined, but we incorporated a decision analysis, based on the comparison of the trees. The decisions sequences and their dynamic are compared using the *nested distance* concept, introduced by Pflug (2010), that captures the structure of the tree and the values of the revealed information. In Pflug and Pichler (2016) the concept is used to compare uncertainty trees, but we will use it to compare the decisions trees. One of the main advantages of this method, is that it keeps the dynamic of the information and the probability of the events and the subsequent paths.

In this chapter, first, we describe briefly the original case, secondly we introduce the notation, third, the variants that lead to the flexibility forms at the same time that the models are presented. Then, we discussed the metrics used in this thesis. Finally, results and discussion are presented.

**Nomenclature**

$\omega \in \Omega$: a specific scenario or leaf of the whole set of scenarios of the tree

$\Omega$: the set of scenarios or leaves.

$\Omega_t'$: the set of scenarios or leaves that present state equal to one at time $t$.

$\Omega_g$: set of scenarios in node $g \in \mathcal{G}$.

$g \in \mathcal{G}$: set of nodes

$\mathcal{G}_t$: set of nodes in stage $t\,:\,(\mathcal{G}_t \subset \mathcal{G}), t \in \mathcal{T}$.

$\omega_g \in \Omega_g$: set of scenarios in node $g \,\forall\, g \in \mathcal{G}$

$\tau_{0,1}$: transition factor between two consecutive rainy periods

$\phi \in [0,1]$: skill level of labor force

$\hat{\beta}_m$: nominal productivity for the resource $m$. If there is only one type, the sub-index is avoided.

$\beta^{\omega}_{tm}$: effective worker $m \in \mathcal{W}$ productivity at time $t \in \mathcal{T}$ in scenario $\omega \in \Omega$ (kilograms per worker per period).

$\breve{\beta}_{t,m}$: actual deterministic productivity for the $m$ resource at moment $t$

$\xi \in \Xi : \{0, 1\}$: the set of possible values that may take the uncertainty realization.

$\bar{\xi}_t$: expected realization in period $t$ (–) $t \in \mathcal{T}$.

$\dot{x}$: decorator for variables (i.e. $x$) that are decided before uncertainty of the period is realizated

$h^{\omega}_{jt}$: daily harvested quantity at $j \in \mathcal{J}$ block in period $t \in \mathcal{T}$ in scenario $\omega \in \Omega$, calculated as $\beta^{\omega}_t z^{\omega}_{jt}$ (kilograms/day).

$x^{\omega}_{tm}$: workers $m \in \mathcal{W}$ hired at time $t \in \mathcal{T}$ (workers) for path $\omega \in \Omega$ (when EV problem, $\Omega = \{1\}$, and belongs to the expected value of the uncertainty).

$y_{tm}$: workers $m \in \mathcal{W}$ laid off at time $t \in \mathcal{T}$ (workers) for path $\omega \in \Omega$ (when EV problem, $\Omega = \{1\}$, and belongs to the expected value of the uncertainty).

$\dot{z}_{jtm}$: workers $m \in \mathcal{W}$ allocated in block $j \in \mathcal{J}$ in period $t \in \mathcal{T}$ before uncertainty happens (workers) for path $\omega \in \Omega$ (when EV problem, $\Omega = \{1\}$, and belongs to the expected value of the uncertainty).

$z_{jtm}$: workers $m \in \mathcal{W}$ allocated in block $j \in \mathcal{J}$ in period $t \in \mathcal{T}$ after uncertainty is revealed (workers) for path $\omega \in \Omega$ (when EV problem, $\Omega = \{1\}$, and belongs to the expected value of the uncertainty).

$l^{\omega}_{mt}$: manpower or labor force $m \in \mathcal{W}$ at time $t \in \mathcal{T}$ (workers) for path $\omega \in \Omega$ (when EV problem, $\Omega = \{1\}$, and belongs to the expected value of the uncertainty).

$\mathcal{T}$: set of stages in the time horizon.

$j \in \mathcal{J}$: a specific block $j$ of the set of blocks of the vineyard.

$m \in \mathcal{W}$: a specific manpower resourcer $m$ of the complete set. If there is only one type, the subindex is avoided.

$r_{01}$: probability rain for two consecutives periods when first period is dry and the second is rainy

$r_{11}$: probability rain for two consecutives periods when both periods are rainy

$w^{\omega}$: conditional probablity for the specific scenario $\omega$

$a_j, b_j, c_j$: quality parameters for the cuadratic equation that represents the quality of the grape in the block $j$

$u_j$: fractional quality loss per rainy period for the grape in block $j$

$B_j$: price of the grape in lot $j$ ($/kilograms).

$C_{E,m}$: cost of hiring ($/worker).

$C_{F,m}$: cost to lay off ($/worker).

$C_{P,m}$: cost of keeping labor idle between periods ($/worker per period).

$C_{H.m}$: cost of harvesting ($/kilograms).

$C_{Z,m}$: cost of assignment before uncertainty is revealed by worker $m \in \mathcal{W}$ ($/kilograms).

$C_{\dot{Z},m}$: cost of assignment after uncertainty is revealed by worker $m \in \mathcal{W}$ ($/kilograms).

$K$: maximum daily reception capacity of the winery (kilograms/day).

$S_j$ $j \in \mathcal{J}$: initial amount of grapes in lot $j$ (kilograms).

$Q_{jt}^{\omega}$: daily quality of the wine grape at $j \in \mathcal{J}$ block in period $t \in \mathcal{T}$ in scenario $\omega \in \Omega$ (–).

$\bar{Q}_{jt}$: average quality for that block $j \in \mathcal{J}$, $t \in \mathcal{T}$.

$\breve{Q}_{jt}$: actual deterministic quality for the block $j$ at moment $t$

$\mathbb{U}_{\mathcal{M}}$: expected value of the solution of a model, i.e., stochastic one.

$\mathcal{M}$: represents any model ant it is useful to write general expressions

$\mathbb{I}_{\mathcal{M}}$: actual income as a percentage of the maximum feasible income for the model $\mathcal{M}$

$\mathbb{L}_{\mathcal{M}}$: cost of labor as a percentage of the maximum income

$\mathbb{Q}_{\mathcal{M}}$: percentage deviation of grape quality from optimum conditions

$\mathbb{S}_{\mathcal{M}}$: percentage of unharvested grapes

$\xi_t^{\omega}$: the value of the uncertainty realization at moment $t$ for the specific scenario $\omega$

$t \in \mathcal{T}$: specific period time in the time span

rk: rookie labor force

ex: expert labor force

$\hat{\beta}_{ex}$: expert individual nominal productivity

$\hat{\beta}_{rk}$: rookie individual nominal productivity

$\theta_{\beta}$: Expert/rookie nominal productivity ratio

$\theta_c$: Expert/rookie cost ratio

$\mathbb{M}$: total expected manpower requirement

$\theta_m$: expected percentage of experts manpower

$\boldsymbol{d}(i,j)$: the value of the nested distance algorithm where $i-th$ is the id number of the first tree to compare, and $j-th$ the id number of the second. It could be described shortly as $\boldsymbol{d}$

$\boldsymbol{d}(\tau)$: having fixed the two trees to be compared, it means the nested distance at any moment between them at time $t=\tau$, with $t \leq \mathcal{T}$

$\mathbb{T}(I), \mathbb{T}(J)$: treed that represent different probability spaces

$i_t$: number of scenarios for tree $\mathbb{T}(I)$, at moment $t$

$j_t$: number of scenarios for tree $\mathbb{T}(J)$, at moment $t$

$n_{i,t} \in \mathcal{N}(\mathbb{T}(I), t)$: the sets of nodes in tree $\mathbb{T}(I)$ at time $t$

$n_{j,t} \in \mathcal{N}(\mathbb{T}(J), t)$: the sets of nodes in tree $\mathbb{T}(J)$ at time $t$

$\hat{\boldsymbol{d}}(n_{i,t}, n_{j,t})$: the value of the Wasserstein distance between two specific distributions

Note 1: when there is a unique type of resource available, the sub index $m$ is avoided.

Note 2: when there is a unique scenario (i.e., deterministic model), the supra index $\omega$ is avoided.

### 4.1. A brief description of the case

In Chapter 3 the model focused on labor hiring, lay-off and allocation decisions in order to maximize the value of the wine grape field. We assumed that a farmer can harvest different types of wine grapes that are divided in lots or blocks. Each block has a specific type of grape, with specific market price, sensitivity to rain effects, and ripening curve. The ripening curve was modeled following a parabola shape; while more narrow this curve, higher quality, and better market price. The optimum epoch for harvesting is when the quality is maximum, and it is a function of time and rain. Rain is the uncertain event, represented by a binary variable, leading to a binomial tree, where probabilities changes according to a Markov two-stages chain. The harvesting task is developed in a time span that lasts 10 days approximately. The blocks are monitored during all the maturity process, that takes different duration (i.e.,60 to 120 days), but in operational terms, the interest in the harvesting task begins when the maturity riches a threshold, and this is the reason for the number of periods that we considered here. The quality is affected negatively by

| Model | Decisions Made at the First Epoch | Decisions Made at the Second Epoch |
|---|---|---|
| MB | $x^{\omega}_{mt}$, $y^{\omega}_{mt}$, $z^{\omega}_{mjt}$ | |
| MA | $x^{\omega}_{mt}$, $y^{\omega}_{mt}$ | $\dot{z}^{\omega}_{mjt}$ |
| MC | $x^{\omega}_{mt}$, $y^{\omega}_{mt}$, $z^{\omega}_{mjt}$ | $\dot{z}^{\omega}_{mjt}$ |

Table 4.1. Different decisions and their epoch in three MS models

the rain. The productivity of the resources is affected by rain also and the risk of future rainfalls could affect the decisions pattern. When the system adds flexibility, the decisions weigh diminishes, but the costs increase, so the balance is something that farmer would prefer. In order to understand this value creation, we study different alternative of system parameter and models, that we call generally, configurations.

### 4.2. Models

Now, we present three formulations of the multistage stochastic problem. The difference among them is the right instant of the resources allocation decision. We will use the word *epoch* to avoid any confusion with the statistical meaning of word *moment*. Decisions made *before uncertainty is revealed in a specific stage*, are made in the *first epoch* and decisions made *after uncertainty realization*, are made in *second epoch*, and both happens in the same period of time. Epoch must not to be confused with stage; in a same stage occurs two epochs. Hiring and firing processes are decisions that must be made at the first epoch, and the allocation decisions are made depending on the approach, according to:

- MS with assignment in the first epoch, MB
- MS with assignment in the second epoch, MA
- MS with assignment in the first epoch with a later correction in the second epoch, MC

Being $x^{\omega}_{mt}$ the hiring decision, $y^{\omega}_{mt}$, the termination one, $z^{\omega}_{mjt}$, allocation decision in the first epoch and $\dot{z}^{\omega}_{mjt}$, allocation decision in the second epoch, the models differences are shown in table 4.1.

The recourse actions are present in MA and MC.

To choose these variants we based on market information. Worker transport could be an extra cost in some cases, i.e. when distances are important, transport could be even a problem of lodging if the task extends over time. Beyond this, to know the probable dates of allocation gives the opportunity to negotiate with local workers instead of considering transportation. Finally, the decision process requires time and meetings, what it is complex in the harvest season, while the efforts are orientated to the coordination of the primary supply chain.

To solve the MS models, we use the *deterministic equivalent model* representation, as in Chapter 3.

The first model to present is MB, model 2.

$$\max \quad \mathbb{U}_{MB} = \sum_{\omega \in \Omega} w^{\omega} \sum_{j \in \mathcal{J}} \sum_{t \in \mathcal{T}} \sum_{m \in \mathcal{W}} \{(B_j Q^{\omega}_{jt} - C_{H,m}) h^{\omega}_{mjt} - C_{E,m} x^{\omega}_{mt} - C_{Z,m} z^{\omega}_{mjt} - C_{F,m} y^{\omega}_{mt} - C_{P,m} l^{\omega}_{mt}\}$$

$s.t.$

$$l^{\omega}_{mt} = l^{\omega}_{m,t-1} + x^{\omega}_{mt} - y^{\omega}_{mt} \quad \forall t \in \mathcal{T},\, \omega \in \Omega, m \in \mathcal{W} \tag{b1}$$

$$\sum_{j \in \mathcal{J}} z^{\omega}_{mjt} \leq l^{\omega}_{mt} \quad \forall t \in \mathcal{T},\, \omega \in \Omega,\, m \in \mathcal{W} \tag{b2}$$

$$\sum_{j \in \mathcal{J}} \sum_{m \in \mathcal{W}} \beta^{\omega}_{mt} z^{\omega}_{mjt} \leq K \quad \forall j \in \mathcal{J}, t \in \mathcal{T},\, \omega \in \Omega, m \in \mathcal{W} \tag{b3}$$

$$\sum_{m \in \mathcal{W}} \sum_{\tau=1}^{t} \beta^{\omega}_{m\tau} z^{\omega}_{mj\tau} \leq S_j \quad \forall j \in \mathcal{J}, t \in \mathcal{T},\, \omega \in \Omega,\, m \in \mathcal{W} \tag{b4}$$

$$x^{\omega}_{mt} = x^{\omega'}_{mt} \quad \forall\, m \in \mathcal{W},\, \omega', \omega\, \in \Omega_g,\, \omega' \neq \omega, g \in \mathcal{G}_{t-1}, t \in \mathcal{T} \,:\, t \geq 2 \tag{b5a}$$

$$x^{\omega}_{m,1} = x^{\omega'}_{m,1} \quad \forall\, \omega', \omega\, \in \Omega,\, \omega' \neq \omega,\, m \in \mathcal{W} \tag{b5b}$$

$$y^{\omega}_{mt} = y^{\omega'}_{mt} \quad \forall\, m \in \mathcal{W},\, \omega', \omega\, \in \Omega_g,\, \omega' \neq \omega, g \in \mathcal{G}_{t-1}, t \in \mathcal{T} \,:\, t \geq 2 \tag{b6a}$$

$$y^{\omega}_{m,1} = y^{\omega'}_{m,1} \quad \forall\, \omega', \omega\, \in \Omega,\, \omega' \neq \omega,\, m \in \mathcal{W} \tag{b6b}$$

$$z^{\omega}_{mjt} = z^{\omega'}_{mjt} \quad \forall\, m \in \mathcal{W},\, j \in J,\, \omega', \omega\, \in \Omega_g,\, \omega' \neq \omega, g \in \mathcal{G}_{t-1}, t \in \mathcal{T} \,:\, t \geq 2 \tag{b7a}$$

$$z^{\omega}_{m,j,1} = z^{\omega'}_{m,j,1} \quad \forall\, \omega', \omega\, \in \Omega,\, \omega' \neq \omega,\, m \in \mathcal{W} \tag{b7b}$$

$$x^{\omega}_{mt}, y^{\omega}_{mt} \in \mathbb{Z}_{+} \quad \forall t \in \mathcal{T},\, \omega \in \Omega,\, m \in \mathcal{W} \tag{b8}$$

$$z^{\omega}_{mjt} \in \mathbb{Z}_{+} \quad \forall t \in \mathcal{T},\, j \in J\,, \omega \in \Omega,\, m \in \mathcal{W} \tag{b9}$$

Model 2. MB model

In this case **all the decisions are made before uncertainty is revealed** or at the first epoch. Anticipating next sections, the manpower linked terms are under-indicated considering the possible set of resources type.

The goal is the maximization of the expected value of the benefits. Income is a function of price, quality, quantity and cost of harvesting, and the costs are connected mainly with labor actions. Constraint $(b1)$ is the balance of manpower, $(b2)$ gives the opportunity to have idle manpower in some periods without the necessity of firing them (if costs are convenient), $(b3)$ is the dairy winery capacity that bounds the dairy harvested total (there is not stocks of harvested grapes because of the acceleration of quality degradation in that context), and $(b4)$ means that the block stock is always greater than zero or zero. The *nonanticipativity principle*, NAC, is represented by $(b5a)$, $(b5b)$, $(b6a)$, $(b6b)$, $(bra)$ and $(b7b)$ constraints (to go deeper in this kind of formulation, see Chapter 3). Constraints $(b8)$ and $(b9)$ are about the nature of the variables.

The second model is MA, presented in model 3.

$$\max \quad \mathbb{U}_{MA} = \sum_{\omega\in\Omega} w^{\omega} \sum_{j\in\mathcal{J}} \sum_{t\in\mathcal{T}} \sum_{m\in\mathcal{W}} \{(B_j Q^{\omega}_{jt} - C_{H,m}) h^{\omega}_{mjt} - C_{E,m} x^{\omega}_{t,m} - C_{\dot{Z},m} \dot{z}^{\omega}_{mjt} - C_{F,m} y^{\omega}_{t,m} - C_{P,m} l^{\omega}_{t,m}\}$$

$s.t.$

$$\sum_{j\in\mathcal{J}} \dot{z}^{\omega}_{mjt} \leq l^{\omega}_{mt} \qquad \forall t\in\mathcal{T},\, \omega\in\Omega,\, m\in\mathcal{W} \qquad (a2)$$

$$\sum_{j\in\mathcal{J}} \sum_{m\in\mathcal{W}} \beta^{\omega}_{t,m} \dot{z}^{\omega}_{mjt} \leq K \qquad \forall j\in\mathcal{J}, t\in\mathcal{T},\, \omega\in\Omega, m\in\mathcal{W} \qquad (a3)$$

$$\sum_{m\in\mathcal{W}} \sum_{\tau=1}^{t} \beta^{\omega}_{\tau,m} \dot{z}^{\omega}_{mj\tau} \leq S_j \qquad \forall j\in\mathcal{J}, t\in\mathcal{T},\, \omega\in\Omega,\, m\in\mathcal{W} \qquad (a4)$$

$$\dot{z}^{\omega}_{mjt} = \dot{z}^{\omega'}_{mjt} \qquad \forall\, m\in\mathcal{W},\, \omega', \omega\, \in\Omega_g,\, \omega'\neq\omega, g\in\mathcal{G}_t, t\in\mathcal{T}, j\in J \qquad (a7)$$

$$\dot{z}^{\omega}_{mjt} \in \mathbb{Z}_{+} \qquad \forall t\in\mathcal{T},\, j\in J\,, \omega\in\Omega,\, m\in\mathcal{W} \qquad (a9)$$

Model 3. MA model

In order to avoid repetition, we indicate the changes compared to MB model. In the objective function, the assignment cost terms considers that decision is made after uncertainty is revealed, changing the variable decorator and the cost of decision. Other important change is the nonanticipativity constraint about allocation. For each uncertainty realization, decision will be fitted to the tree that begins in the next node. Constraint $(a7)$ replaces $(b7a)$ and $(b7b)$. The other changes are easy to follow.

MC approach considers variables $z$ and $\dot{z}$, to refer both before and after correction. In this model, the cost of making the decision before is less than making the decision after, and the values are similar to those used in MB and MA. In this work $C_{Z,m} < C_{\dot{Z},m}$. The MC model is model 4 and presents only the changes compared to model 2.

$$\max \quad \mathbb{U}_{MC} = \sum_{\omega\in\Omega} w^{\omega} \sum_{j\in\mathcal{J}} \sum_{t\in\mathcal{T}} \sum_{m\in\mathcal{W}} \{(B_j Q^{\omega}_{jt} - C_{H,m}) h^{\omega}_{mjt} - C_{E,m} x^{\omega}_{mt} - C_{Z,m} z^{\omega}_{mjt} - C_{\dot{Z},m} \dot{z}^{\omega}_{mjt} - C_{F,m} y^{\omega}_{mt} - C_{P,m} l^{\omega}_{mt}\}$$

$s.t.$

$$l^{\omega}_{mt} = l^{\omega}_{m,t-1} + x^{\omega}_{mt} - y^{\omega}_{mt} \qquad \forall t\in\mathcal{T},\, \omega\in\Omega, m\in\mathcal{W} \qquad (c1)$$

$$\sum_{j\in\mathcal{J}} (z^{\omega}_{mjt} + \dot{z}^{\omega}_{mjt}) \le l^{\omega}_{mt} \qquad \forall t\in\mathcal{T},\, \omega\in\Omega,\, m\in\mathcal{W} \qquad (c2)$$

$$\sum_{j\in\mathcal{J}} \sum_{m\in\mathcal{W}} \beta^{\omega}_{mt} (z^{\omega}_{mjt} + \dot{z}^{\omega}_{mjt}) \le K \qquad \forall j\in\mathcal{J}, t\in\mathcal{T},\, \omega\in\Omega, m\in\mathcal{W} \qquad (c3)$$

$$\sum_{m\in\mathcal{W}} \sum_{\tau=1}^{t} \beta^{\omega}_{m\tau} (z^{\omega}_{mj\tau} + \dot{z}^{\omega}_{mj\tau}) \le S_j \qquad \forall j\in\mathcal{J}, t\in\mathcal{T},\, \omega\in\Omega,\, m\in\mathcal{W} \qquad (c4)$$

$$x^{\omega}_{mt} = x^{\omega'}_{mt} \qquad \forall\, m\in\mathcal{W},\, \omega', \omega\, \in\Omega_g,\, \omega'\ne\omega, g\in\mathcal{G}_{t-1}, t\in\mathcal{T} \,:\, t\ge 2 \qquad (c5a)$$

$$x^{\omega}_{m,1} = x^{\omega'}_{m,1} \qquad \forall\, \omega', \omega\, \in\Omega,\, \omega'\ne\omega,\, m\in\mathcal{W} \qquad (c5b)$$

$$y^{\omega}_{mt} = y^{\omega'}_{mt} \qquad \forall\, m\in\mathcal{W},\, \omega', \omega\, \in\Omega_g,\, \omega'\ne\omega, g\in\mathcal{G}_{t-1}, t\in\mathcal{T} \,:\, t\ge 2 \qquad (c6a)$$

$$y^{\omega}_{m,1} = y^{\omega'}_{m,1} \qquad \forall\, \omega', \omega\, \in\Omega,\, \omega'\ne\omega,\, m\in\mathcal{W} \qquad (c6b)$$

$$z^{\omega}_{mjt} = z^{\omega'}_{mjt} \qquad \forall\, m\in\mathcal{W},\, j\in J,\, \omega', \omega\, \in\Omega_g,\, \omega'\ne\omega, g\in\mathcal{G}_{t-1}, t\in\mathcal{T} \,:\, t\ge 2 \qquad (c7a)$$

$$z^{\omega}_{m,j,1} = z^{\omega'}_{m,j,1} \qquad \forall\, \omega', \omega\, \in\Omega,\, \omega'\ne\omega,\, m\in\mathcal{W} \qquad (c7b)$$

$$\dot{z}^{\omega}_{mjt} = \dot{z}^{\omega'}_{mjt} \qquad \forall\, m\in\mathcal{W},\, \omega', \omega\, \in\Omega_g,\, \omega'\ne\omega, g\in\mathcal{G}_{t}, t\in\mathcal{T}, j\in J \qquad (c8)$$

$$x^{\omega}_{mt}, y^{\omega}_{mt} \in \mathbb{Z}_+ \qquad \forall t\in\mathcal{T},\, \omega\in\Omega,\, m\in\mathcal{W} \qquad (c9)$$

$$z^{\omega}_{mjt},\, \dot{z}^{\omega}_{mjt} \in \mathbb{Z}_+ \qquad \forall t\in\mathcal{T},\, j\in J\,, \omega\in\Omega,\, m\in\mathcal{W} \qquad (c10)$$

Model 4. MC model

### 4.2.1. Expected Value Model, EV

The EV model is much like a deterministic model where the known event is replaced by its expected value taking into account the distribution. Quality and uncertainty expected

value are calculated as was previously introduced in Chapter 3. As EV model considers that the future is known perfectly (the expected scenario is taken as the real one), later corrections to decisions policy have no value, because they are not needed essentially. We must consider instead, that making decisions before and after have different cost, so even when the expected events do not change, but due to that cost, the decision may be.
The EV approach was formulated as in the previous chapter. To illustrate, we present the EA model (model 5), because EB is similar to the presented in Chapter 3. Model EC does not exist in actual terms; giving the opportunity to change a decision lately does not offer an advantage in context where future is perfectly known.

$$\max \quad \sum_{j\in\mathcal{J}}\sum_{t\in\mathcal{T}}\sum_{m\in\mathcal{W}}\{(B_j\bar{Q}_{jt}-C_{H,m})h_{j,t,m}-C_{E,m}x_{mt}-C_{\dot{Z},m}\dot{z}_{j,t,m}-C_{F,m}y_{mt}-C_{P,m}l_{mt}\}$$

$s.t.$

$$l_{mt}=l_{m,t-1}+x_{mt}-y_{mt} \quad \forall t\in\mathcal{T}, m\in\mathcal{W} \qquad (ea1)$$

$$\sum_{j\in\mathcal{J}}\dot{z}_{mjt}\le l_{mt} \quad \forall t\in\mathcal{T},\, m\in\mathcal{W} \qquad (ea2)$$

$$\sum_{j\in\mathcal{J}}\sum_{m\in\mathcal{W}}\bar{\beta}_{mt}\dot{z}_{mjt}\le K \quad \forall j\in\mathcal{J}, t\in\mathcal{T}, m\in\mathcal{W} \qquad (ea3)$$

$$\sum_{m\in\mathcal{W}}\sum_{\tau=1}^{t}\bar{\beta}_{m\tau}\dot{z}_{mj\tau}\le S_j \quad \forall j\in\mathcal{J}, t\in\mathcal{T},\, m\in\mathcal{W} \qquad (ea4)$$

$$x_{mt}, y_{mt}\in\mathbb{Z}_+ \quad \forall t\in\mathcal{T},\, m\in\mathcal{W} \qquad (ea9)$$

$$\dot{u}_{mjt}\in\mathbb{Z}_+ \quad \forall t\in\mathcal{T},\, j\in J,\, m\in\mathcal{W} \qquad (ea10)$$

Model 5. EA model

We kept the id number of the constraints on the MS model even when the number of them is larger than the require in EA model. The scenarios have been reduced by the expected value of the uncertainty. The NACs are not present because the information is available since the planning step.

In order to compare with an upper bound, we solve models *Wait and See*, WS. A WS model is similar to a MS but the NAC are inactive because we make decisions considering perfect information about future, so each scenario does not need to be limited for other possible events. Similarly to EC model variant, the WC model is only a mathematical issue because with perfect information since the beginning, later corrections are not required. WS approach has two flavors, WB and WA, mainly differentiated by the cost of decision. WB is similar to MB and WA, to MA, but with a relaxation of NACs.

### 4.3. Model Comparison Metrics

To compare the solutions obtained by using different approaches, we will use three well known metrics (Birge and Louveaux, 2011; Escudero et al., 2007):

- the *expected value of perfect information* (EVPI),
- the *value of the stochastic solution* (VSS)
- the *nested distance* ($\boldsymbol{d}$)

The first two were introduced in Chapter 3. The nested distance is a metric that allows the comparison of two trees. We use this concept to consider if there are changes in the decisions tree through the models. The *nested distance*, was introduced by Pflug (2010). In Pflug and Pichler (2016) there is a very extensive content to know better details, especially in the mathematical aspects. In the appendix B we offer a summary and some experiments to understand better the intuition of the metric. Now we will describe how nested distance works.

If there are two trees, we want to establish if they are similar in all their elements: structure, nodes value, and probabilities associated with their arcs. This method proposes to obtain a numerical value whose magnitude indicates the difference between two specific trees. Two trees where their structure, value of nodes and associated probabilities are known will be compared. The first step is to position both trees in their last period, $\mathcal{T}$. Subsequently, for each of the scenarios, a predecessor node from which they originate was identified. Thus, in period $\mathcal{T}-1$ we identify a list of nodes that represent the complete set of predecessors for the final step. Each of those nodes is the predecessor of one or more scenarios, and the total probability of the scenarios that belongs to a specific node, is equal to one.

The first step is to operate both trees, $\mathbb{T}(I)$ and $\mathbb{T}(J)$, to obtain a first matrix of distances between scenarios. So, there are $i_{\mathcal{T}}$ and $j_{\mathcal{T}}$ scenarios for first and second tree respectively, at moment $\mathcal{T}$. The matrix distances is done for all the combinations of scenarios, so at time $\mathcal{T}$ the matrix of distance has a size of $(i_{\mathcal{T}} \times j_{\mathcal{T}})$. This matrix is formed by distances among scenarios, taking their absolute value.

The second step, is to gather the scenarios that share a common predecessor node. Being $\mathcal{N}(\mathbb{T}(I), t)$ and $\mathcal{N}(\mathbb{T}(J), t)$ the sets of nodes in first tree and second tree respectively, at time $t = \mathcal{T} - 1$, we go for all $n_{i,t} \in \mathcal{N}(\mathbb{T}(I), t)$ and list the scenarios that are gathered in it and their probability distribution. We do the same for $n_{j,t} \in \mathcal{N}(\mathbb{T}(J), t)$. So for each combination of $(n_{i,t}, n_{j,t})$ we also have a list of probabilities and a list of distances (obtained before). As the distances are similar for both trees, the only difference at that node combination are the probability distributions. The method proposes to reduce this structure through a distance that takes into account the probability distribution. Based on the optimal transport problem, the authors suggest the *Wassertein distance*, $\hat{\boldsymbol{d}}$. So, after two nodes are compared, $n_{i,t}$ and $n_{j,t}$, their differences are reduced to $\hat{\boldsymbol{d}}(n_{i,t}, n_{j,t})$. After going through all the nodes of a time period, we have a collection of differences between nodes. Now, the original distance matrix of size $(i \times j)$ is reduced to a matrix of size $(n_{i,t} \times n_{j,t})$. This matrix considers that the actual size of the trees is given by the number of nodes in period $t$, because the rest of the branches has been reduced to that size.

This procedure goes on until there is no more stages, and the trees are reduced to a final distance, the nested distance, $\boldsymbol{d}(i, j)$. The nested distance method is a generalization of the Wasserstein distance, where Euclidean distance are used. As nested distance summarizes the structure, node information and probabilities. It is difficult to understand the specific and individual impact of the elements on the final magnitude. Experiments in appendix B help in this sense. From those, we reinforce some behaviors, considering two trees and their relative results:

- The nested distance increases according to the differences between the two trees that appeared earlier.

- While the node information value is big, so was the nested distance.
- The probability distribution impacts more on the final nested distance if the node that is most likely has a different magnitude than those nodes with which it is compared. In another term, the difference between two nodes will be expressed more if this combination is supported by a greater probability of occurrence.

Vitali (2018) indicates that the objective function value in a multi-stage stochastic problem is positively correlated to the nested distance, but there is no information about more concrete relationships. They go even further in Horejšová et al. (2020), and the conclusions are in the same line. We think that there is a space for an exhaustive study of the numeric behavior of nested distance, but it is out of this thesis scope. Finally, for measure instances, $\boldsymbol{d}$ has been proved in the literature valid and meaningful in relatives analysis.
The implementation of the algorithm for $\boldsymbol{d}$ was made in Python 3.0, using a python library to calculate the WD, called POT (Flamary et al., 2021), where the algorithm approximates the Sinkhorn instead of $\hat{\boldsymbol{d}}$ as it is described by Cuturi (2013).

### 4.4. Results and Discussion

Results will be presented in two parts. The first is about the value created by the decision epoch as a flexibility source. The second is on the contribution of different harvesting resources in this context. In the first part, we compared the profit performances of the models for several conditions of rain and resource ability. The main aim is to understand if the epoch of the decisions created value and in what degree. We also studied how quality patterns affect the models' contributions. At the end of this part, we compared the decisions policies through $\boldsymbol{d}$, in order to explore if decisions policies are really different. The second part is about the contribution of teams instead of any unique type of resource. First, we explained the experiments after which the model updates, and then, the results are produced. A cost-sensitive analysis was developed to understand how the staff changes to ensure the highest expected profit value.
The parameters of models and the quality patterns are similar to the used in Chapter 3. The models were implemented using Python, written for PYOMO Python (Hart et al., 2017) and the optimization engine was GUROBI v. 8.1.0. The solution time was not limited, the

optimality gap was set to 1% and the integrality parameter was the solver default. We used a laptop computer with an Intel Core processor i7-6700HQ CPU 2.60 GHz, with 32.0 GB of RAM memory running Windows 10.

We tested a total of 405 instances per model and decision epoch, so the total number of optimization running was 2835 considering the variants of MS (3), EV (2), and WS (2). The number of instances is the combination of five rain probabilities ($r_{01} = (0.1, 0.3, 0.5, 0.7, 0.9)$), three rain probability transition factor, three levels of flexibility ($\phi = (0.3, 0.5, 0.7)$), three ripening pattern, three available combinations of resources (rookies, experts, both). For the cost sensitivity we added more than 600 optimization runs.

### 4.4.1. Decision epoch and contribution to the value of the system

Farmer makes three general decisions, hiring, lay-off and allocation the resources to different blocks. As allocation decision is being evaluating as a source of flexibility, we will present its effects.

MA is a variant of the MS that considers information about uncertainty occurrence in order to make the correct assignment. Two derived questions are: if the epoch of making the assignment decision changes, how much does it impact in the expected value of the system?; and, what are the conditions that affect this value?

To explore these questions, we compare the expected profit value of MS in identical conditions for MA, MB and MC. MC acts like a kind of option: we can make a decision before uncertainty happens assuming first cost, that is cheaper than to make the same decision after uncertainty realization. If that decision is not so useful for the conditions of the system once uncertainty reveals, decision could be corrected paying an extra cost. In Figure 4.1, we compare $\mathbb{U}_{MC}$ and $\mathbb{U}_{MA}$ to $\mathbb{U}_{MB}$ as a ratio. MB was selected to be the reference because it is the most disadvantageous model.

Chart $(a)$ shows the relative behavior for MA to MB. For low rain probabilities, regardless of the resource ability, MB seems to be better than MA, even when differences are very little. This makes sense because the decisions in MA are more expensive than in MB, and in context where risk is not so important, an extra cost is not necessary. The MA value

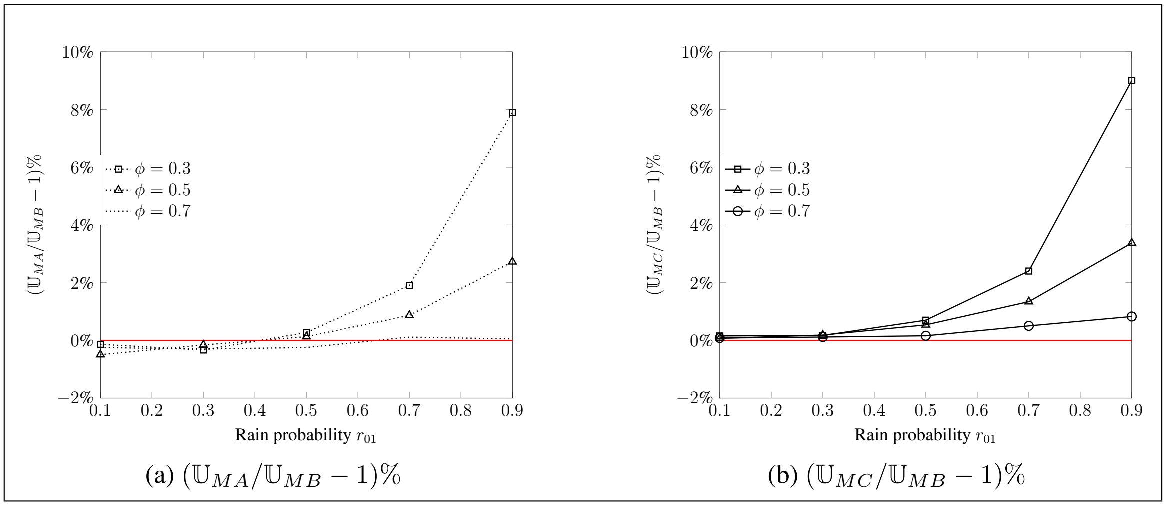


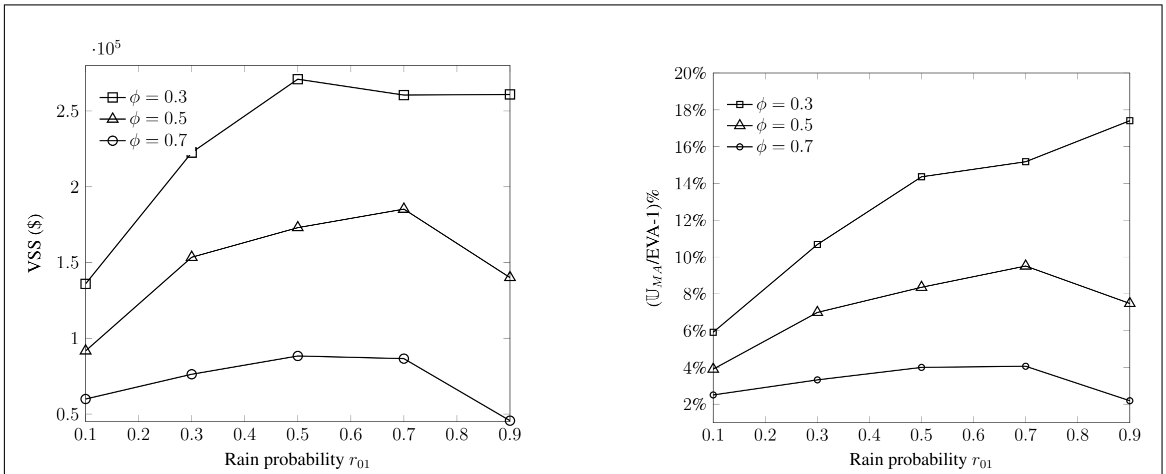

Figure 4.1. MA and MC performances compared to MB, expressed as the ratio of their economic expected final result. Three levels of skill are examined in the harvesting resource, under the SOD ripening pattern.

increases when rain conditions are more probable. When decisions could be made after uncertainty happens in risky environment it has more value for the system. Flexibility in resources, helps as it was studied in Chapter 3.

Chart $(b)$ shows the relative performances for MC and MB. MC is the more flexible model, because it gives the opportunity to decide at any epoch. For low rain probabilities, MB and MC are similar in performance. In this case it is supposed that MC decision policy is similar to the MB and the cost are similar too. When rain probability increases MC is more advantageous than MB and even than MA. In a very extreme condition of rain it is interesting to see that MC is even better than MA; this comes from the nature of decisions. There are some decisions that could be made before uncertainty happens, and if they require fine tuning, the correction could be made.

To highlight is the efficiency of models around maximum uncertainty. As the three variants are MS models, there is not significant contribution in that situation of the epoch of assignment.

In terms of information requirement, when rain probability is less than 40%, the refresh of the system is not highly appreciated. To invest in the update of the uncertainty information, seems to be more important when the risk of bad performances in farm is present. The possibility of having fresh information captures value, and the gap growth is exponential.




(a) VSS=$\mathbb{U}_{MA}$-EVA (b) ($\mathbb{U}_{MA}$/EVA-1)%

Figure 4.2. VSS considering EVA and MA $(a)$, for three levels of flexible rookie resource, under the SOD ripening pattern. In chart $(b)$ the results are shown in a relative way.

### 4.4.2. Flexible resources influence in decision epoch

As we discussed in Chapter 3, the flexibility of the resource compete with the contribution of the model. Changing the epoch of decision, the impact of flexible resources could be less important, so even for low flexible resources results would be suitable. Let's see how this intuitive though is mathematically discovered.

To begin, we discuss the MA model, which is a traditional recourse model. Figure 4.2, chart $(a)$, shows the VSS value for MA and EA approaches (in that case, we denoted by EVA) for different conditions of uncertainty, represented by the rain probability.

When $\phi$ is near 1, the contribution of the MA model is poor, so the difference between both approaches, could be negligible. If the skill level decreases, the VSS increases, giving to MA more value because of its advantage of considering the uncertainty in an exhaustive way, and more general in an explicit form. The shape of the VSS curve is a bell considering the rain probability. If $r_{01}$ is very low, the weight of the scenarios will be concentrated, and EV approach captures in a good manner this situation. When rain event is highly probable, EV consider this as a fact and the behavior is like a worst case. This vision lacks some of the positive scenarios that should be used to leverage better results, something that explains the gap with MA. The expected value of the rain event, will be similar to the

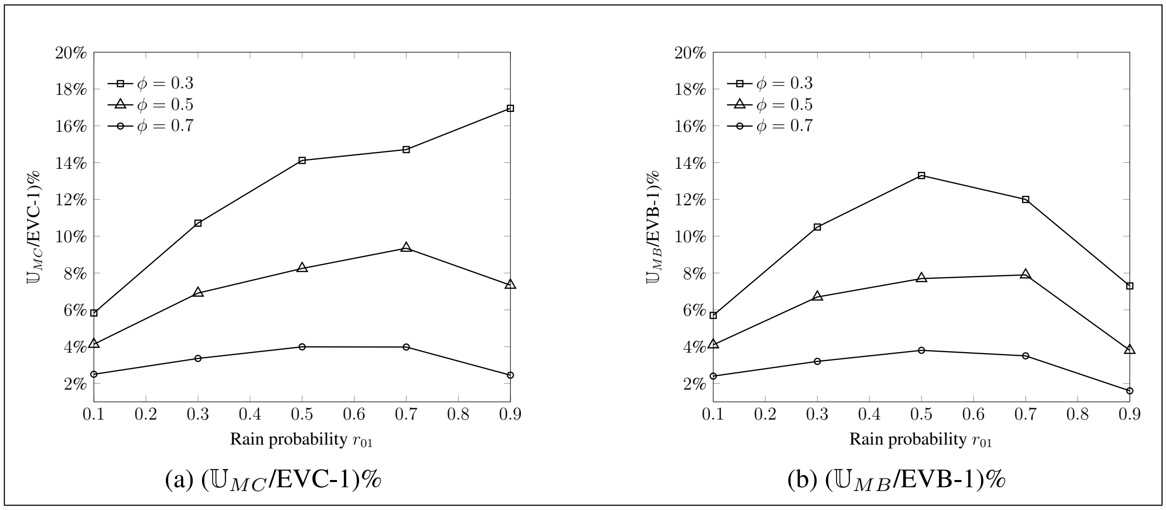



Figure 4.3. Relative Performance in terms of expected value of the economic profit for MC and EVC $(a)$, and MB and EVB $(b)$, for three levels of skill in the rookie resource, under the SOD ripening pattern.

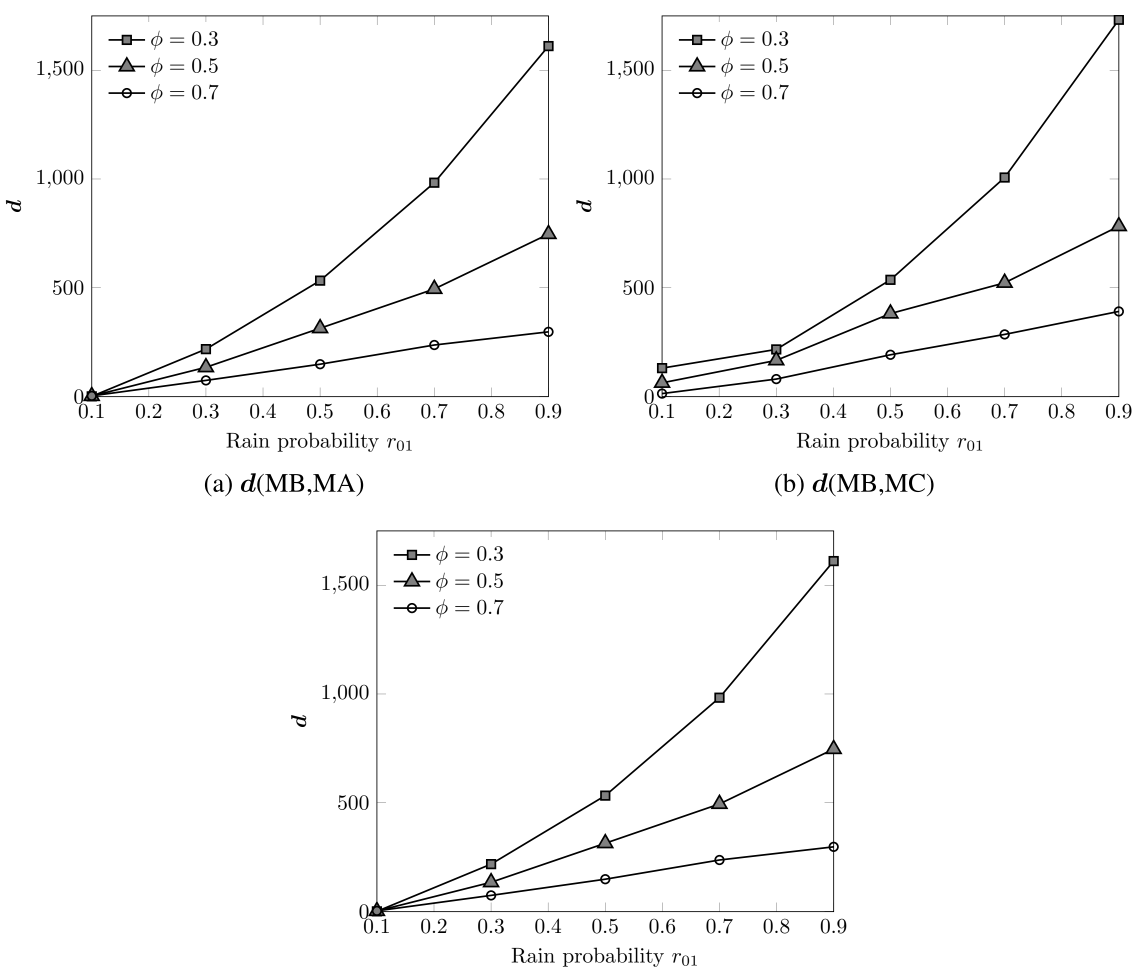
heavier scenarios in the extremes. When uncertainty increases, that is with $r_{01}$ around 0.5, there is no preference for the occurrence of specific scenarios, so the performance of the EA gets worse and MA is a good choice.

In Figure 4.2, chart $(b)$, we observed performance ratio. The shape of the curve keeps similar to the shown chart 4.2-$(b)$. It is interesting to remark, that for poor skills, the relative advantage of MA reaches levels greater than 18%.

Now we extend the flexible resource impact to the other models, especially the relative analysis. In Figure 4.3, we see the relative behavior for the MB and MC models considering EVB (the EB solution value) and EVC (the EB solution value, similar to EC), respectively. In the first graph $(a)$, the shape is alike to MA variant. For both decisions moments, MS approach creates value, but it was more restricted when decisions need to be made before uncertainty realization, $(b)$ chart.

### 4.4.3. Comparing decisions sets

As time goes by, the decisions that the manager makes could be different between the support models. There are three types of decisions: hiring, terminations, and allocation or assignment. Only the last one offers some flexibility to the decision-maker according to the type of the model, EV, MA, MB or MC. MS models are different to EV solutions because




(a) $\boldsymbol{d}$(MB,MA)

(b) $\boldsymbol{d}$(MB,MC)

(c) $\boldsymbol{d}$(MA,MC)

Figure 4.4. Final Nested Distance for labor decision tree in different conditions of uncertainty and resources flexibility - Models MA, MB, and MC (SOP ripening pattern)

of the nature of the assumptions. But it is not clear yet if the MS alternative solutions are very different between them. In this section, we will concentrate in comparing decisions sets for MS variants. Decisions should be different, if the expected value of the system is. To advise about those differences we study how the trees of decisions differ among them through the nested distance concept. Firstly, we explore the MS variants final gaps, and secondly, the behavior of nested distance over time.

In Figure 4.4 there are three charts that present $\boldsymbol{d}$(MB,MA), $\boldsymbol{d}$(MB,MC) and $\boldsymbol{d}$(MA,MC). The decisions trees are about labor stock, a way of summarizing the hiring and termination decisions. The three charts present similar patterns. Low abilities increments the nested distance value, according rain probability is higher. MB and MA presents differences even at low rain probability; the value is similar to the observed when MB and MC are compared, probably because set of decisions and risk even when low could require some adjustment after uncertainty is revealed. MA and MC, are very similar when rain probability is low, but as it increases, the decisions set are also more diverse.

In Figure 4.5 we see the $\boldsymbol{d}$(MB,MA), $\boldsymbol{d}$(MB,MC) and $\boldsymbol{d}$(MA,MC), for three different lots, two flexibility levels and different rain probabilities.

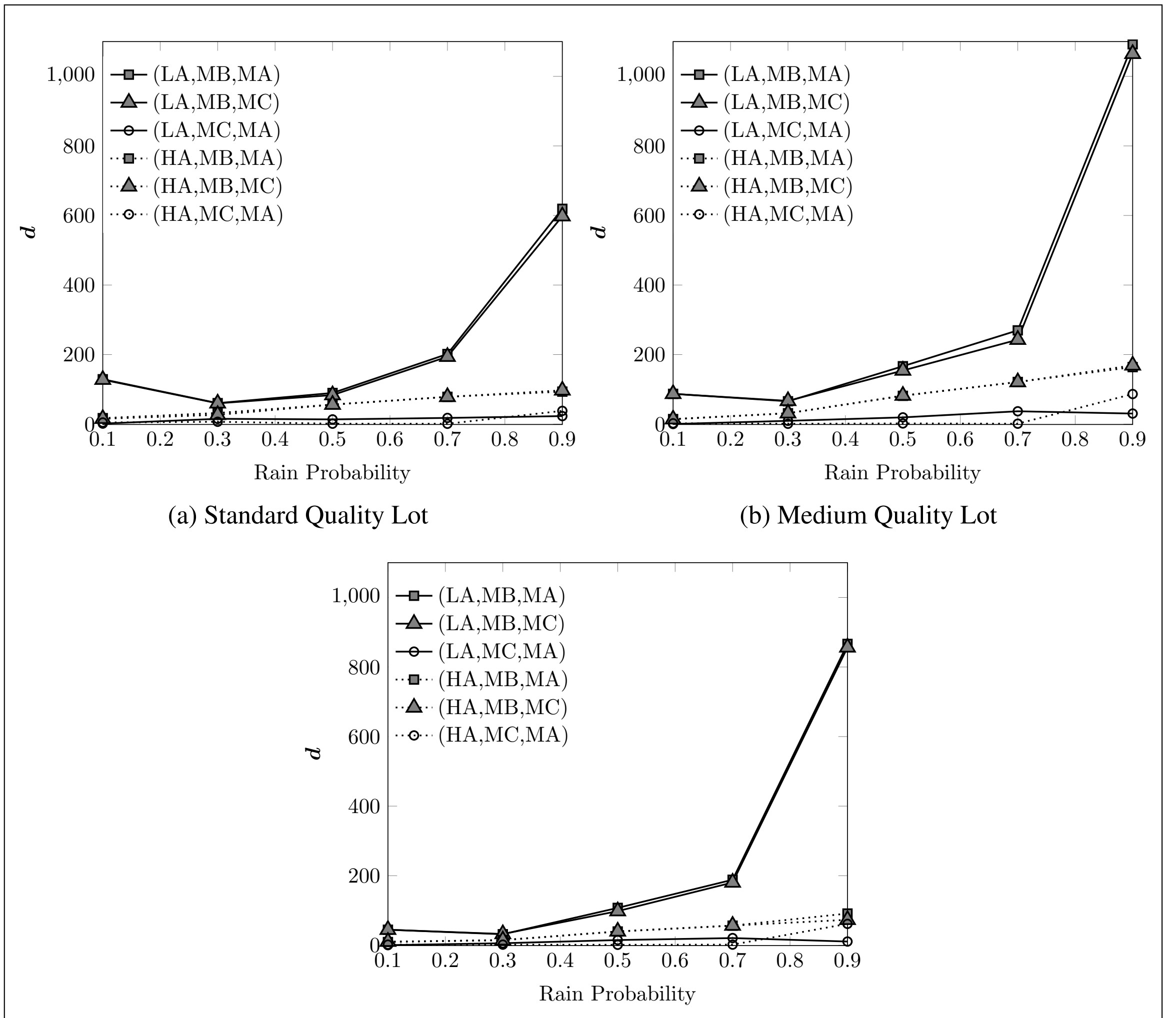
In the standard quality lot (where time and rain have low impact), high ability decreases the nested distance among the models. In context of high ability, MA and MC present low differences in the trees. When ability is low, the distances between trees rise quickly when $0.5 \leq r_{01}$. This pattern has been observed previously in labor trees distances.

The second chart and third chart in Figure 4.5, keep the same pattern than first chart, previously explained, but for medium and high quality respectively. All the charts show that MA and MC are very similar, something that is reinforced by $\boldsymbol{d}$(MB,MC) $\approx$ $\boldsymbol{d}$(MA,MC).

The $\boldsymbol{d}$ parameter fits well to the expected behaviors of the decisions policies, with the advantage of summarizing this intuition in just a single metric. Beyond the total nested distance, it is interesting to see the rate of change, in order to understand if there are specific moments in which the decisions trees diverge strongly. According to Figure 4.5, when $r_{01} = 0.9$ and $\phi = 0.3$, the decisions sets are very different. We will review that situation for the three lots and for the labor decisions trees.

Figure 4.6 indicates two cases that present great divergence, both when rain probability is high, and resources ability are low and high, respectively. For low flexible resources, the differences are present from the beginning. The $\boldsymbol{d}$(MB,MA) and $\boldsymbol{d}$(MB,MC) behave linearly and in a very similar path. The $\boldsymbol{d}$(MA,MC) is very small according fig 4.5; however, the divergence is more pronounced in the first 7 periods than in the last.

When ability is very high, we know that differences trends to be minor between models. Even in that case, there is an evolution of the nested distance value that changes over




(a) Standard Quality Lot

(b) Medium Quality Lot

(c) High Quality Lot

Figure 4.5. Final Nested Distance for different conditions of uncertainty and resources flexibility for three different lots in the allocation decision-Models MA, MB, and MC (SOP ripening pattern)

time. Chart (b) in fig 4.6 shows the evolution of $\boldsymbol{d}$ for high ability resources. $\boldsymbol{d}$(MB,MA) indicates that in the first five periods the decisions about labor forces are quite similar. Just when the optimum day has passed, decisions set changes, because MA offers later corrections that are useful in this context. We could appreciate the effect of the high flexible resource with this delay; the decisions divergences could wait because there is a strong capability to react. $\boldsymbol{d}$(MB,MC) shows a progressive distance increment; period five is the inflection point, accelerating the differences because MC model reacts using its later correction capability. Finally, $\boldsymbol{d}$(MA,MC) evolution is at constant rates. Remember that

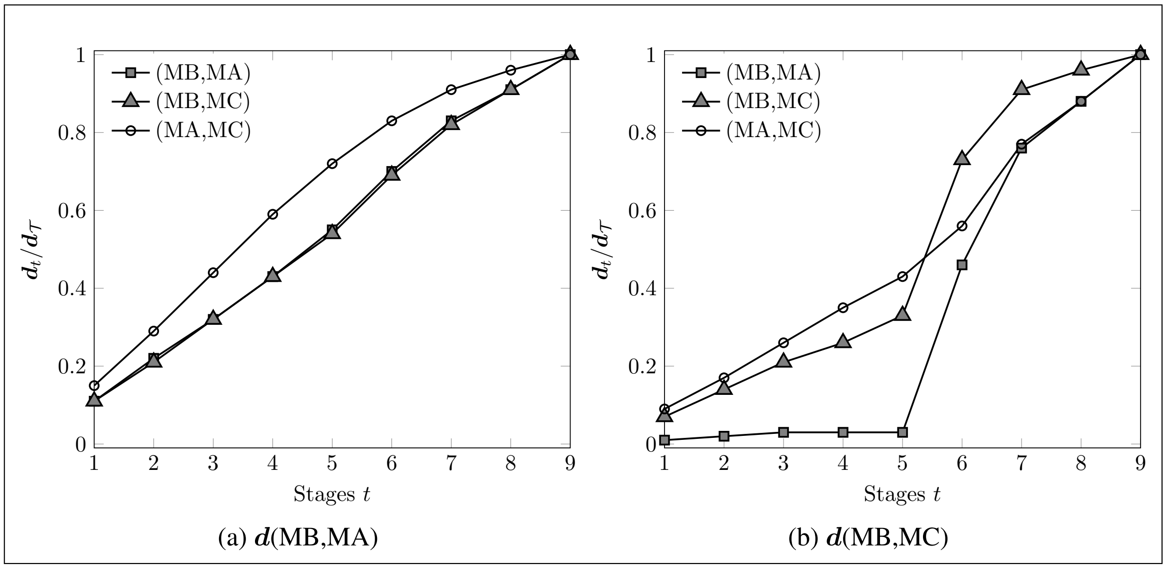


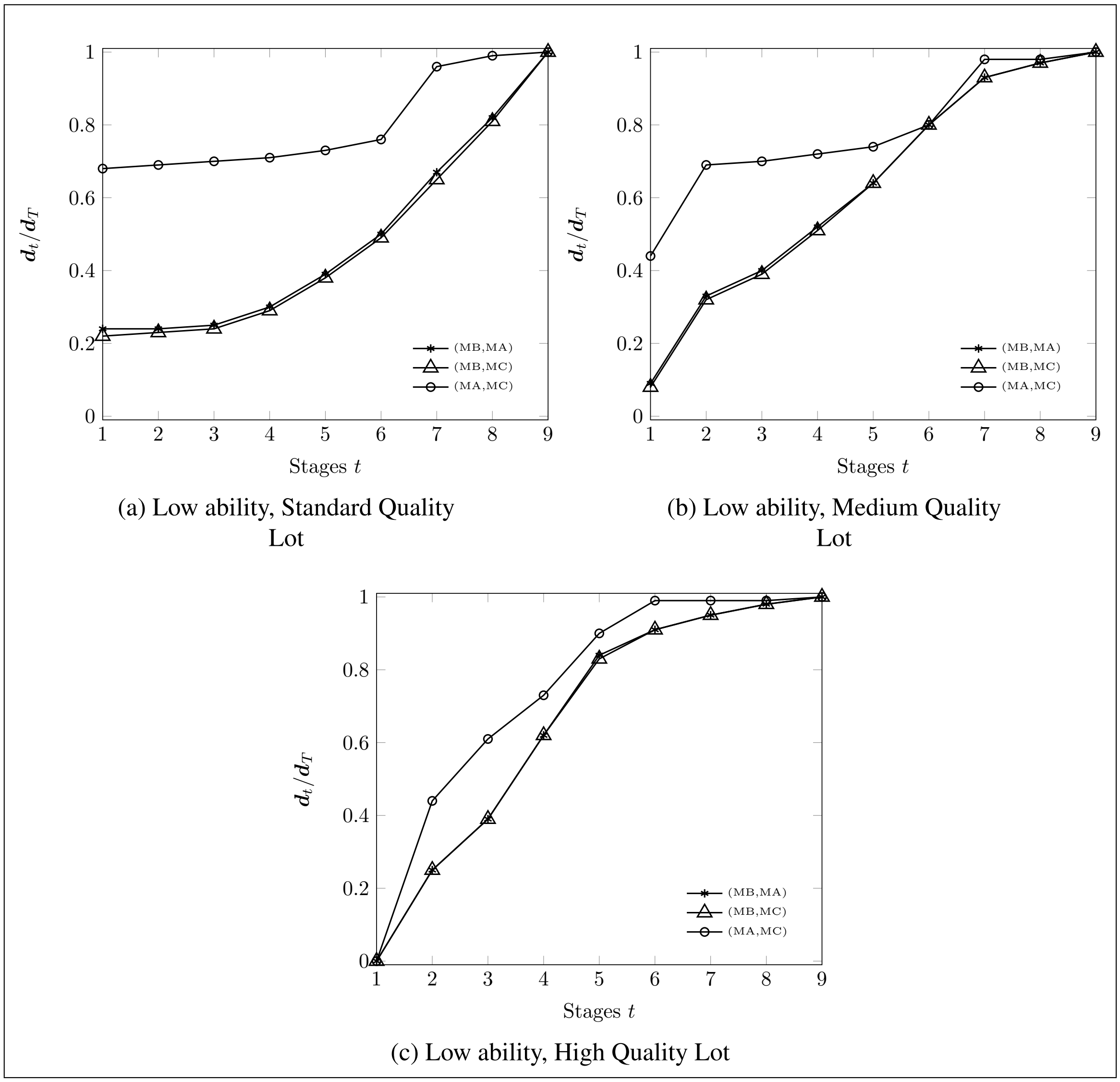

Figure 4.6. Nested Distance Evolution for labor decision set when rain probability is high. Chart (a) is for $\phi = 0.3$ and chart (b) is for $\phi = 0.7$ (SOP ripening pattern)

even when final distances could be very short, there is an evolution that may be used to indicate inflection points.

Let us see what happens with the evolution of nested distance over time when allocation decisions are studied.

In Figure 4.8 there are three charts; each of them belongs to a specific type of grape, differentiated by quality. In chart (a), $\boldsymbol{d}$(MB,MC) and $\boldsymbol{d}$(MB,MA) are very similar for the first lot. Both distances say that (MB,MC) and (MB,MA) have been different from the beginning. This divergence is attenuated in the first period but it increases linearly. For $\boldsymbol{d}$(MA,MC) there is a big difference at the beginning. MC is more versatile than MA; MA makes decisions that are more expensive even when they are similar to MC. In this case, MC differs from the policy from the beginning. it is fundamental to remember that the final magnitude of $\boldsymbol{d}$(MA,MC) is independent of what we see in the evolution analysis. Charts (b) and (c) smooth the evolution speed, linearizing it.

In Figure 4.8, chart (d), when resource is highly skilled, all the models show very similar behaviour until period five. The features of each models are used in the last part of the harvesting period. $\boldsymbol{d}$(MB,MC) and $\boldsymbol{d}$(MB,MA) follows the same evolution path. Even




(a) Low ability, Standard Quality Lot

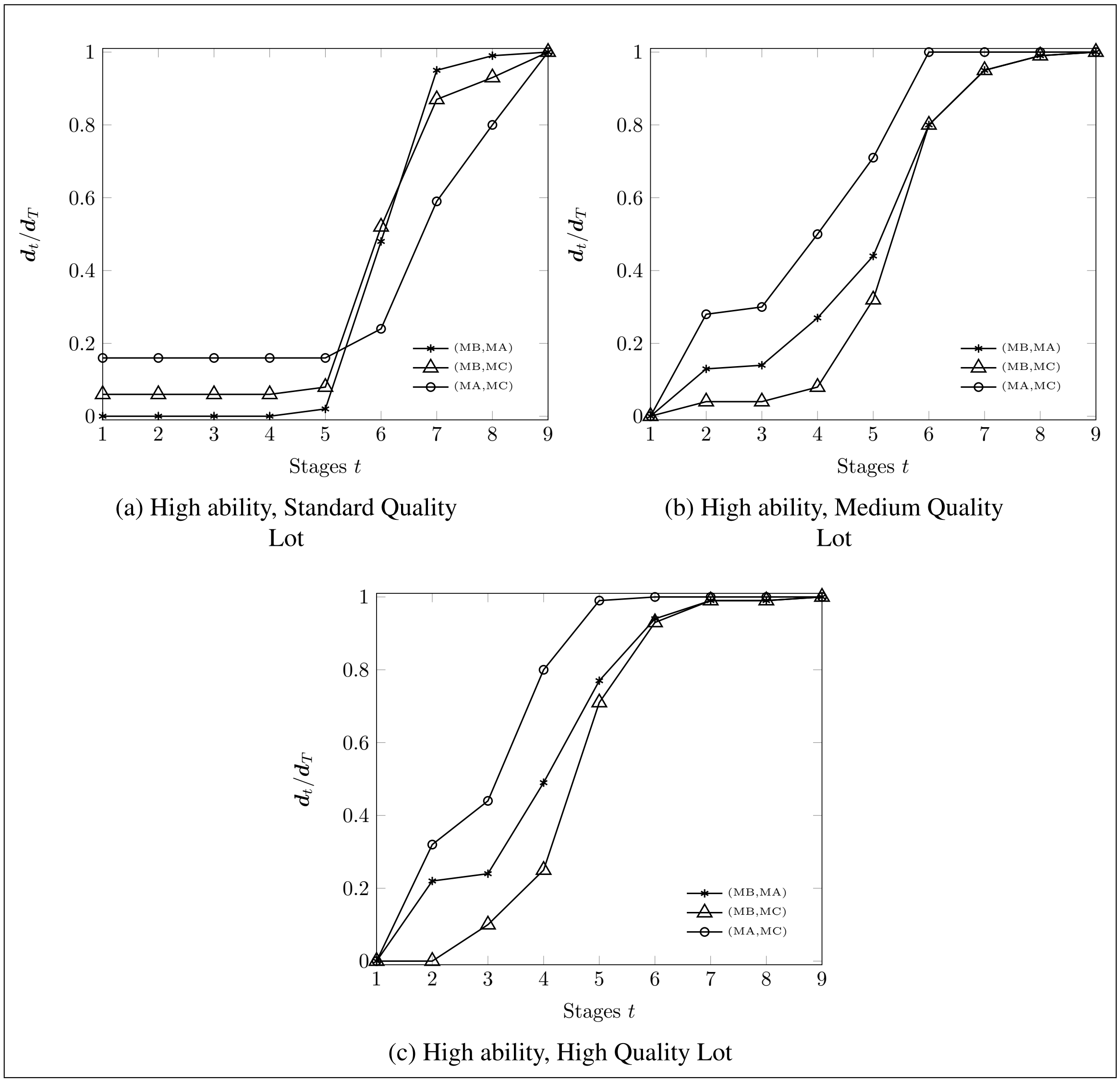
(b) Low ability, Medium Quality Lot

(c) Low ability, High Quality Lot

Figure 4.7. Evolution of the nested distance between (MB,MA), (MB,MC) and (MA,MC) allocation trees when rain probability is high for SOP ripening pattern - PART A

when in magnitude $\boldsymbol{d}$(MA,MC) is small, the evolution pattern is similar to the previous ones.

In chart (e) the change rate is smoothed and more distributed in all the periods. In chart (f), there is an interesting pattern. In this case, decisions differs in the first five periods, and beyond that, the divergence is completely developed. This conduct coincides with the idea of harvesting in advance because of the high sensibility of the grape to the rain effects.




(a) High ability, Standard Quality Lot

(b) High ability, Medium Quality Lot

(c) High ability, High Quality Lot

Figure 4.8. Evolution of the nested distance between (MB,MA), (MB,MC) and (MA,MC) allocation trees when rain probability is high for SOP ripening pattern - PART B

### 4.4.4. Resources Flexibility: Harvest Resources Team

As we mentioned previously, the flexibility value is connected with the cost and the benefit of using it. Benefit is a consequence of the occurrence of the event for which the flexibility was thought. Also, the benefit depends on the extension of the event that faces flexibility; in some cases the extension should be very important to give the opportunity to generate value throw flexibility execution.

In this section we study the effects of having the opportunity to hire different skilled resources, to constitute a team. The problem offers the opportunity to contract two types of resources that differ in the nominal productivity keeping similar the skill to work under rainy conditions. Rookies have less productivity than experts, but their contracts are cheaper too. Farmer must decide if more expensive resource is useful, and then, makes the hiring decisions. There is sensitivity to the price of the resource, because as cost increases, the possibility of repayment is smaller.

To conduct this exploratory study we have reduced the number of experiments trying to diminish the moving parts. Block number was reduced to one, with 900,000 kg of initial stock, and one type of grape at a time. Grapes quality are standard and premium, where the latter's price is three times the first one. The set of resources is $W = \{rk, ex\}$, rookie and experts, respectively. Expert/rookie nominal productivity ratio, $\theta_\beta$ is used to formulate the nominal productivity of the expert, $\hat{\beta}_{ex} = \theta_\beta \hat{\beta}_{rk}$. There is also a cost ratio, $\theta_c$, that allows to express the hiring cost of the expert as $C_{E,ex} = \theta_c C_{E,rk}$. This is the unique cost that changes for experts and rookies. Both ratios are added to the model, to represent the new resource. For example, MC's objective function will be:

$$\max \sum_{\omega \in \Omega} w^\omega \sum_{j \in \mathcal{J}} \sum_{t \in \mathcal{T}} \sum_{m \in \mathcal{W}} \{(B_j Q^\omega_{jt} - C_{H,m}) h^\omega_{mjt} - C_{E,rk} x^\omega_{rk,t} - \theta_c C_{E,rk} x^\omega_{ex,t}$$
$$- C_{Z,m} z^\omega_{mjt} - C_{\dot{Z},m} \dot{z}^\omega_{mjt} - C_{F,m} y^\omega_{mt} - C_{P,m} l^\omega_{mt}\} \quad (4.1)$$

The harvested quantity in a period $h^\omega_{mjt}$ could be written in a general way as $\sum_m \beta^\omega_{mt}(z^\omega_{mjt} + \dot{z}^\omega_{mjt})$; explicitly, $\beta^\omega_{rk,t}(z^\omega_{rk,jt} + \dot{z}^\omega_{rk,jt}) + \theta_\beta \beta^\omega_{rk}(z^\omega_{ex,jt} + \dot{z}^\omega_{ex,jt})$

Each leaf of the tree will have a total requirement of manpower during the harvest process, and also a weight to be pondered. We define the expected number of total manpower requirement, expressed by equation 4.2. To consider the distribution of the resources in the team, we defined the $\theta_m$ index as the ratio of expert manpower to total expected requirement (equation 4.3).

$$\mathbb{M} = \sum_{\omega \in \Omega} w^{\omega} \sum_{t \in \mathcal{T}} (l^{\omega}_{ex,t} + l^{\omega}_{rk,t}) \tag{4.2}$$

$$\theta_m = \frac{\sum\limits_{\omega \in \Omega} w^{\omega} \sum\limits_{t \in \mathcal{T}} l^{\omega}_{ex,t}}{\sum\limits_{\omega \in \Omega} w^{\omega} \sum\limits_{t \in \mathcal{T}} (l^{\omega}_{ex,t} + l^{\omega}_{rk,t})} = \frac{\sum\limits_{\omega \in \Omega} w^{\omega} \sum\limits_{t \in \mathcal{T}} l^{\omega}_{ex,t}}{\mathbb{M}} \tag{4.3}$$

The intuition says that low prices for expert resource (similar to the rookies one) will lean towards to their hiring. In the same way, if experts are extremely expensive, their number will be very low. In the middle, there are several combinations, where decisions are not linear and support is required. We explore that zone. Before continuing, in Table 5.1, the rest of the parameters values to be tested in this study are shown.

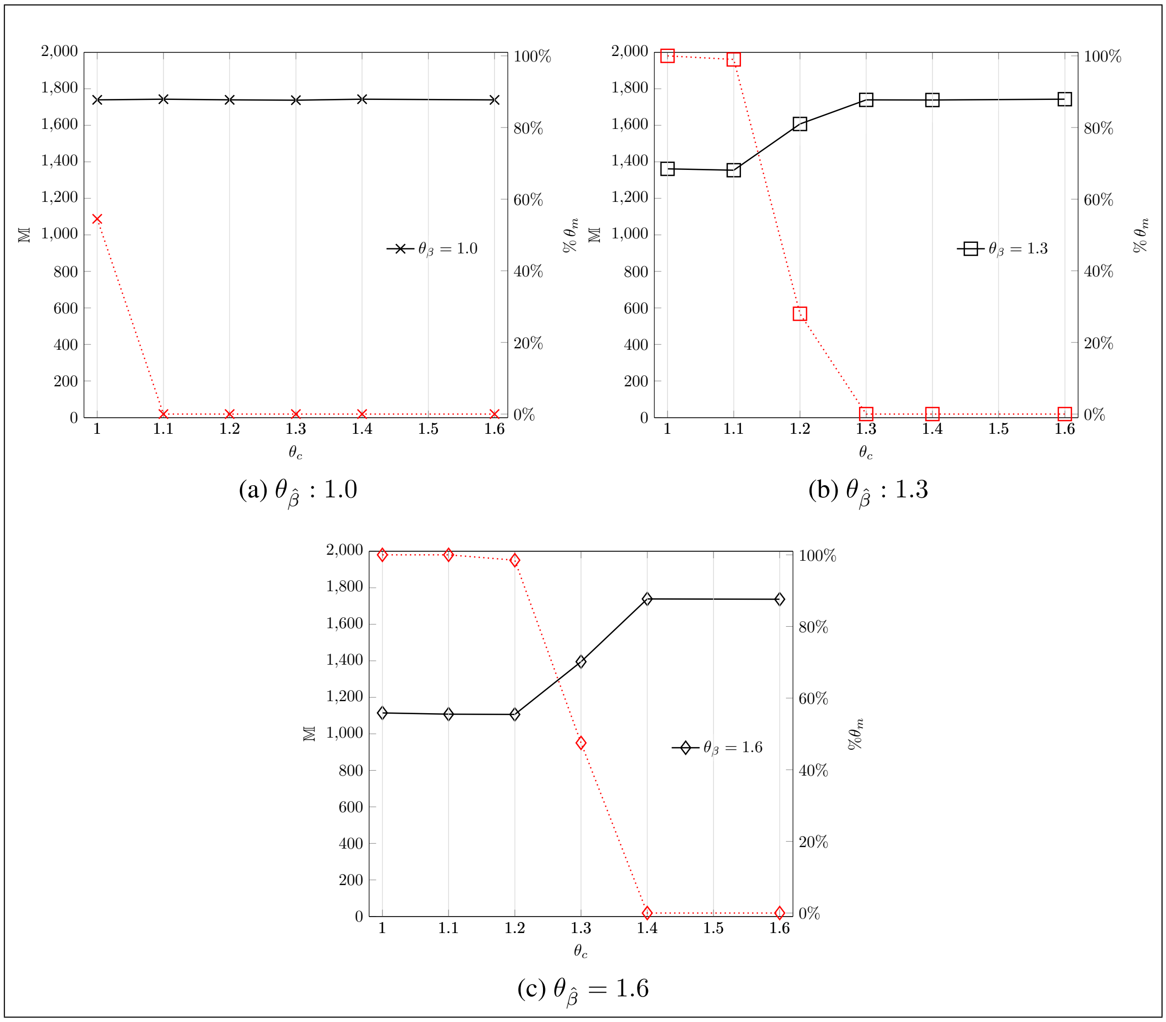
Table 4.2. Parameters for the different experiments

| Feature | Notation | Values | Units |
|---|---|---|---|
| Workers Ability | $\phi$ | 0.3, 0.7 | – |
| Expert/rookie nominal productivity ratio | $\theta_\beta$ | 1.0, 1.15, 1.30, 1.45, 1.60, 1.75 | – |
| Expert/rookie cost ratio | $\theta_c$ | 1.0, 1.05, 1.10, 1.15 | – |
| | | 1.20, 1.30, 1.40, 1.60, 1.80 | – |
| Rain probability | $r_{01}$ | 0.1, 0.5, 0.9 | – |
| Transition factor | $\tau_{11}$ | 1.0 | – |
| Rain quality penalty | $u_j$ | 2% | % |

In Figure 4.9 we present $\theta_m$ and $\mathbb{M}$ for different $\theta_c$ for EV and MA. We returned to these cases because there are two possible and basis approaches. The total workers requirement is on the left y-axis, expert percentage is on the right y-axis and $\theta_c$ is on x-axis. The skill level is $\phi = 0.3$, grape is standard and $r_{01} = 0.5$. The dashed line values is read in the right y-axis, and the solid line values is read in the left y-axis.

In Figure 4.9 there are three charts that offer a lot of information.

The x axis is the expert/rookie cost ratio; as it increases, expert resource is more expensive. In the left y-axis the expected total manpower requirement is represented, $\mathbb{M}$; in the right y-axis, we present the percentage of the total expected manpower that is expert, $\theta_m$. We




(a) $\theta_{\hat{\beta}}$ : 1.0

(b) $\theta_{\hat{\beta}}$ : 1.3

(c) $\theta_{\hat{\beta}} = 1.6$

Figure 4.9. Total manpower (expressed as total of hired days) (left y-axis, solid line), $\theta_m$ (right y-axis, dotted line), for different conditions of costs, $\theta_c$ (x-axis), for productivity expert/rookie ratio, $\theta_\beta : (1.0, 1.3, 1.6)$, when skills are low, **standard quality** and **rain probability is around 0.5**

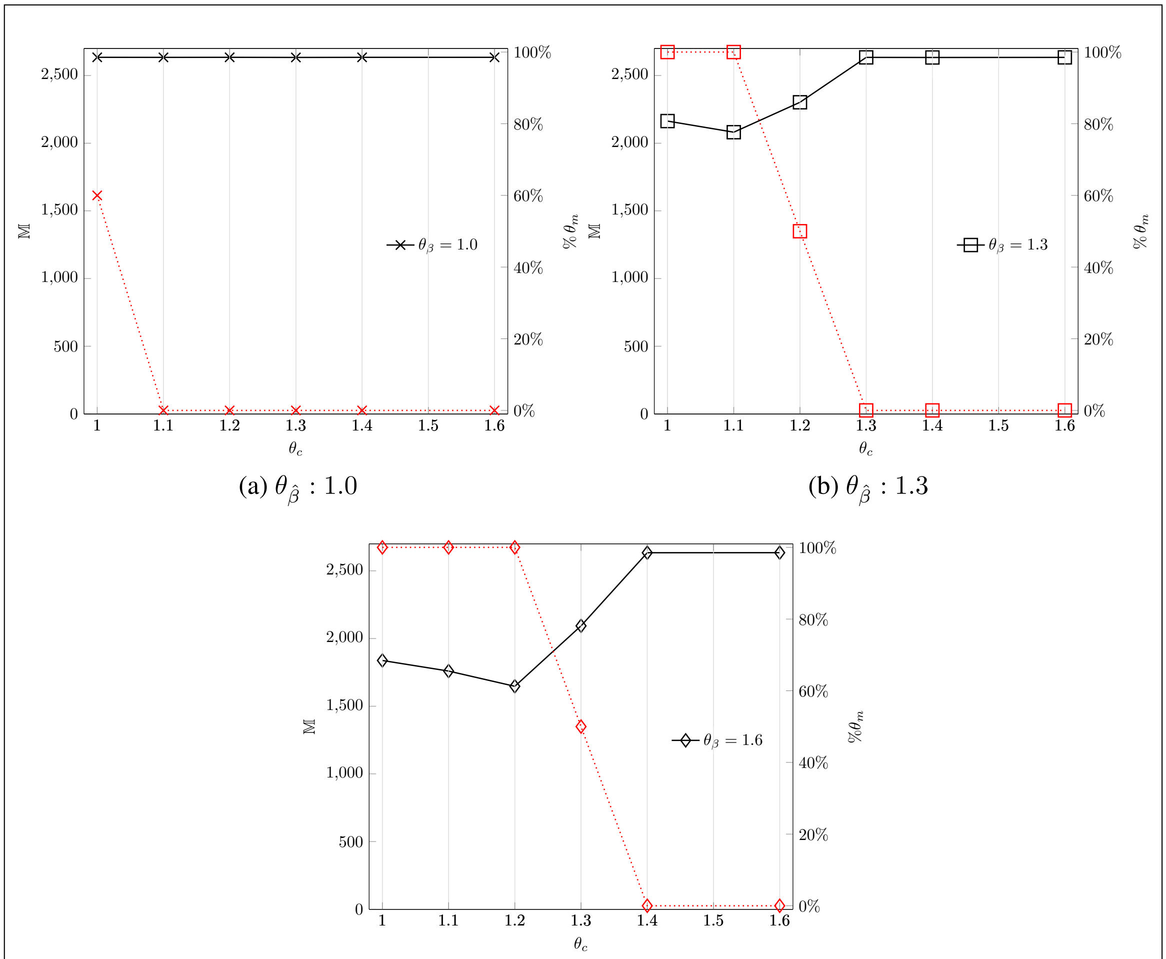
set in this Figure $r_{01} = 0.5$, harvesting **standard quality**. Each of the three charts offer the sensitivity results considering different productivity ratios for expert/rookie. In chart $(a)$, the productivity of expert is similar to the rookie's one, $\theta_\beta = 1.0$. At the same cost, $\theta_c = 1.0$, the decision is not relevant because the resources behave in a similar way, including their costs. When $\theta_c > 1.0$ but $\theta_\beta = 1.0$, then the ratio $\theta_c/\theta_\beta$ increases meaning that expert is more expensive than rookie even when the productivity is the same. In this case the expected requirement of experts decays to zero, and rookies are chosen like harvesting resources. Further this point, the team is made up of only rookies.

Chart $(b)$ offers a higher level of productivity for the expert, $\theta_\beta = 1.3$, $30\%$ more than rookie. For ratios $\theta_c/\theta_\beta < 1$, the productivity unit is cheaper than the rookie option. At low costs, the team is completely integrated by experts; according cost increases, rookies are an option, mixing the resources. It is interesting to highlight that for ratios like $(1.2/1.3)$ the option of expert is weaker. These decisions are an effect of flexibility; although experts are cheaper than rookie in term of nominal productivity unit price, they imply a bigger cost that may be hard to recover. The opportunity to recover the investment is the weak spot in the some flexibility analysis, because the real value is linked to the use of that flexibility (Rogalski, 2011). In chart $(c)$, the value feature of flexibility is strongly present. In this case, $\theta_\beta = 1.6$ and the choice of expert, according to an economic concept, is reasonable in the range $\theta_c = [1, 1.6)$, because in that interval $\theta_c/\theta_\beta < 1$. In spite of that expected behavior, expert resource is an alternative in the range $\theta_c = [1, 1.4)$. The reason is that the cost is too high to bear and recovery is not assured. In this context, the expected total manpower, $\mathbb{M}$ increases about 30% when expert resources are not hired; for this particular configuration, the decision based on an average productivity is wrong, so the MS model shows the importance to consider the stochasticity in an explicit way.

When rain probability is high, the expert contribution may change and they can be more valuable for the system. In order to analyze it we present in Figure 4.10 information about team constitution.

In the context of Figure 4.10, in chart $(a)$ the behavior is similar to the case when rain probability is 0.5. Charts $(b)$ and $(c)$ are a little different to the previously discussed. For example, in chart $(c)$, according expert cost increase, the expected total manpower is minor, but it is still completely constituted by experts.

We present the same results in Figure 4.11, but in a compact way. Three charts present the $\theta_\beta : (1.0, 1.3, 1.6)$ cases, and each of them shows the behavior for $r_{01} = 0.5$ and $r_{01} = 0.9$. Again the dynamic of the expert choice is conditioned by the risk. The total expected labor force is bigger than in the case of standard quality, mainly because the features of the product, both in maturity curve and price.




(a) $\theta_{\hat{\beta}}$ : 1.0

(b) $\theta_{\hat{\beta}}$ : 1.3

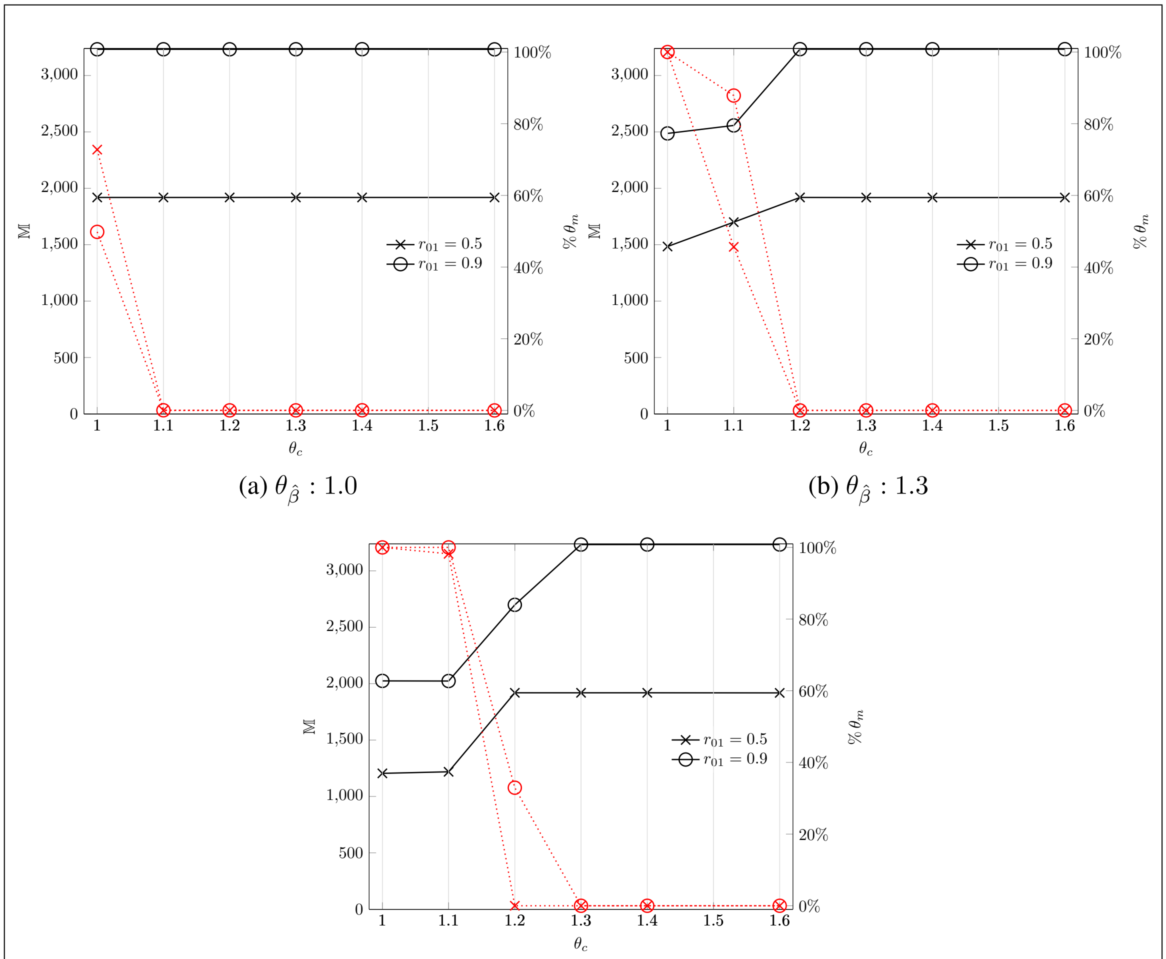
(c) $\theta_{\hat{\beta}} = 1.6$

Figure 4.10. Total manpower (expressed as total of hired days) (left y-axis, solid line), $\theta_m$ (right y-axis, dotted line), for different conditions of costs, $\theta_c$ (x-axis), for productivity expert/rookie ratio, $\theta_\beta : (1.0, 1.3, 1.6)$, when skills are low, **standard quality** and **rain probability is around 0.9**

## 4.5. Conclusions

In this chapter we explore the value of flexibility, considering two main sources: decision epoch and manpower teams. This is an expansion of the work in Chapter 3, because we add a variant with two decision epoch for the same variable (allocation, model MC). About resources flexibility, we add to the rain ability, the nominal productivity as a source, dividing the group of workers in two, experts and rookies.

In MS approach, we observe that the decision epoch creates value, especially when uncertainty is important. The recourse action is not always needed, and the largest values




(a) $\theta_{\hat{\beta}}$ : 1.0

(b) $\theta_{\hat{\beta}}$ : 1.3

(c) $\theta_{\hat{\beta}} = 1.6$

Figure 4.11. Total manpower (expressed as total of hired days) (left y-axis, solid line), $\%, \theta_m$ (right y-axis, dotted line), for different conditions of costs, $\theta_c$ (x-axis), when $\theta_\beta : (1.0, 1.3, 1.6)$, skills are low, **premium quality** and $r_{01} : \{0.5, 0.9\}$

are obtained when decisions are made before with minor costs, and later correction are allowed, model MC. The cost of the later correction and the previous decision affect the flexibility acquisition, so the cost analysis is a must in order to understand the dynamic of the managerial job. In this line, the comparison among different approaches is necessary, because the goodness of different models finally depends on the tolerance of the decision-maker.

As we have seen in previous chapter, models contribution could be less important if resource skills are high. The nominal productivity is a flexibility source, but the effect is not so clear as in the case of ability. Expensive resources ensure quick answers for some scenarios, but in other case they only ensure a higher cost that it is not tackled by the manager. At this point, we see that expert resources are not always preferred even when their cost per unit of productivity may be better than the rookie one. This result could be intuitive, but we indicate that flexibility has value only when a resource is used, and in that context, the initial paid in the hiring process make the resources unprofitable if an MS approach is used, where stochasticity is explicit.

This part of the chapter offers important opportunities to improve the study. Statistical analysis considering the original results for each of the scenarios in the more of thousand variant per models, may offer an interesting shortcut to understand the risk, and main numbers behind the complex MS optimization. Those shortcuts could be of interest in the real-world application.

Finally, we also address the decision analysis through nested distance concept. This distance is a way of summarizing the difference between decisions maps, being a novel application, even when the principle and uses are pretty the same that introduce (Pflug and Pichler, 2016). The $\boldsymbol{d}$ shows a relationship with uncertainty, tree size, model, variant, and values of decisions; probably we can identify how each of them affects the $\boldsymbol{d}$ value, but it is difficult to predict, because the decision map is function of the business opportunity. Results follows some appreciations of the reducing scenarios theory Vitali (2018); Timonina (2015); Horejšová et al. (2020) that agrees on the capacity of $\boldsymbol{d}$ to be used as a relative metric to compare trees, but that correlates weakly with the objective function value. In our application, we note that the divergence between solutions trees is gradual and it is not concentrated in a specific period. However, in the allocation variable, we observed very similar patterns to those observed in the progress of the harvest. The changes in $\boldsymbol{d}$ represent correctly the dynamic and gives an idea of how different are the trees at each point in the timeline. Additionally, the decisions trees become more different when rain probability increases and resources flexibility decreases; in those cases the $\boldsymbol{d}$ for the variant of MS growth exponentially.

# Chapter 5

# Rolling tree Approach

Flexibility is a system attribute that is recognized as a way of handling uncertainties. T. Wang (2005) indicated that flexible thinking is more proactive than the robust approach, but requires better abilities and tools for management. Therefore, we presented mathematical programming models in the previous chapters that serve as part of the quantitative tools. The MS model behaves similar to the management practice, with the number of decisions and when they should be made being two sources of flexibility for the manager Mandelbaum and Buzacott (1990). The MS approach similar to the managing practice could be mathematically intractable because of its size. The discretization process for the probability distribution and the impact of the time span on the explosion of the tree and computational times tend to increase exponentially. Along this line, researchers have been working for decades to develop practices linked to the scenario reduction (refer to Chapter 2 for literature details).

In this chapter, we faced the size problem but with a slightly different vision to the reducing and generating scenarios proposed in literature. The rolling horizon approach uses an algorithm that allows the opportunity to update information according time goes by. This approach is a way of implementing a multi-period model. The decisions are made using an expected value of the uncertainty events; with the plan being periodically revisited to re-optimize it with new available information. In this chapter, we proposed a rolling approach that uses sub-trees keeping part of the original granularity of the information (original tree), instead of expected value simplification. The *rolling tree algorithm*, RTA, is a rolling algorithm where each planning step is solved through a partial tree in the short term, and a simplification of the uncertainty in the long term. This algorithm is smaller than the complete MS model, so the computational effort is minimized; however, the loss of information may have a negative impact on the final decisions. The aim of this chapter is to shed light on: whether RTA can replace the MS approach while still maintaining a controlled loss of performance, which conditions of uncertainty and context RTA compete with the MS approach, and finally, how the temporal parameters affect the value of the RTA as an alternative to the MS?

Before going any further, it is necessary to make an explanation of the information update. In this work, MS considers that the stochastic process leads to all the feasible nodes from the very beginning. It means that when time goes by, the probability of the rain events follows the stochastic process and new information about this stochastic process, i.e. probability changes, does not occur. In practical terms, the farmer has an MS solution, a tree that has all the feasible paths, and according to the history of events, individualized nodes that describe the status of the system and the decisions to be considered in future scenarios. When the rolling horizon approach is used, there is the advantage of added information at the moment of system review, for example, the changes in the rain probabilities. However, in this study, the information was kept constant according to the Markovian process. Hence, what kind of update could be made to the rolling horizon approach in this chapter? We considered that the update to the rolling horizon approach is linked to the possibility of having more detailed information about future, or in the possibility of knowing a specific node. In both cases, the updated process gives more information about the future without changing the data that is available for the MS approach. In this way, the economic performance between MS and rolling horizon approach, as well as the computational times, could be compared.

The rest of the chapter is organized as follows. Initially, we presented the rolling horizon approach. After that, the rolling tree approach was outlined in detail. Subsequently, the computational experiments and metrics were discussed, and the results and discussion presented immediately after. Finally, the main conclusions were drawn and presented.

**Nomenclature**

$k \in \{1 \ldots \mathcal{K}\}$: number of planning process in a rolling scheme

$n \in \{1 \ldots \mathcal{N}\}$: all the nodes that could be the initial node for a planning task

$\mathbb{T}$ , $\mathbb{T}_i$: the very original tree under study; when two trees should be differentiated, a subscript could be used

$t \in \mathcal{T}$: time periods in the whole complete time span $card(\mathcal{T})$

$\xi \in \Xi$: uncertainty set

$\mathcal{P}(\Xi)$, $\mathcal{P}_t(\Xi)$: probability distribution of the set of uncertain events. As distribution could changes over time, a $t$ subscript could be added to indicate that distribution belongs to a specific period

$\bar{\mathcal{P}}_t(\Xi)$, $\bar{\mathcal{P}}$: expected value of the distribution at moment $t$. If there is no chances of confusing, the simplest notation is used, $\bar{\mathcal{P}}$

$\mathbb{T}^k$, $\mathbb{T}^k_i$: the planning step $k$ original tree; if $k = 1$ then the tree is the very original.

$\bar{\mathbb{T}}^k$: the simplified or reduced tree version of $\mathbb{T}^k$

$\mathcal{T}^k$, $(t^k_p : \mathcal{T})$: time span for the planning step $k$, that begins in $t^k_p$ and finishes in $\mathcal{T}$

$\mathcal{I}^k$: initial information for the planning step $k$

$\mathcal{I}^k_t$: information available at period any period $t$ during planning step $k$.

$\mathcal{D}^k$: decisions set that are obtained by optimization in planning step $k$

$\Delta_r$: time span for system review

$t^k_p$: initial period of planning for the cycle $k$

$t^k_r$: revision period in cycle $k$; the status of the system in this point is the initial status for planning cycle $k + 1$.

$\Delta_z$: frozen time span

$t^k_z$: final period of the execution of decisions made in cycle $k$

$t^k_s$: first period for the simplification trees.

$n \in \mathcal{N}$: number of running of the planning process to evaluate the expected value of the approach rolling horizon approach

$U^k_n$: partial profit of the planning step $k$ in the planning process running $n$

$\mathbb{U}_{RH}$: the expected value of the RHA solution. General notation

$\hat{g} \in \hat{\mathcal{G}}$: the list of nodes that could be used as initial nodes for the planning step. These nodes belong to different periods

$\hat{t} \in \hat{\mathcal{T}}$: the list of periods where a planning step begins, in accordance with a previous time structure

$\dot{\mathbb{T}}(k, \hat{g})$: the primary tree that begins in node $\hat{g}$ in the planning step $k$

$\mathcal{I}(k, \hat{g})$: the information for the planning step $k$ in node $g$

$\ddot{\mathbb{T}}(k, \hat{g})$: secondary tree, generated from $\dot{\mathbb{T}}(k, \hat{g})$

$\breve{g} \in \mathcal{G}(k, \hat{g}, t_r^k)$: a node that belongs to the set of terminal nodes of the secondary tree $\ddot{\mathbb{T}}(k, \hat{g})$ when is review at period $t_r^k$

$\dddot{\mathbb{T}}(\breve{g})$: tertiary tree that initiates in node $\breve{g}$

$\hat{\mathbb{T}}(\breve{g})$: means the tertiary tree, $\dddot{\mathbb{T}}(\breve{g})$: after the simplification process

$\hat{\dot{\mathbb{T}}}(k, \hat{g})$: the consolidated primary tree after the reduction process

$t_s^k$: the first period of the simplified part of the secondary tree

$\Delta_f$: time span for the secondary tree fan

$g_0$: initial node for implementing a decision policy

$g_d$: destiny node when a decision policy has been implemented

$\mathcal{F}$: mathematical transformations of information received to conduct from the origin node to the destiny one.

$\mathbb{U}_{RE}$: expected value of the profit in a RTA using EV reduction

$\mathbb{U}_{RW}$: expected value of the profit in a RTA using WC reduction

$U_{g_d}$: profit in node $g_d$ obtained by simulation

$\boldsymbol{d}$: nested distance

$\tau(\hat{\dot{\mathbb{T}}}(k, \hat{g}))$: Optimization Time for the consolidate tree

$\tau_t^\omega$: computational time or effort for the optimization process that begins in time $t$ , scenario $\omega$ , that keeps correspondence with the $\hat{g} \in \hat{\mathcal{G}}$

$\omega \in \Omega$: a specific scenario or leaf of the whole set of scenarios of the tree

$\Omega$: the set of scenarios or leaves.

$\Omega_t'$: the set of scenarios or leaves that present state equal to one at time $t$.

$\Omega_g$: set of scenarios in node $g \in \mathcal{G}$.

$g \in \mathcal{G}$: set of nodes

$\mathcal{G}_t$: set of nodes in stage $t\ :\ (\mathcal{G}_t \subset \mathcal{G})$: $t \in \mathcal{T}$.

$\omega_g \in \Omega_g$: set of scenarios in node $g \,\forall\, g \in \mathcal{G}$

$\tau_{0,1}$: transition factor between two consecutive rainy periods

$\phi \in [0,1]$: skill level of labor force

$\hat{\beta}_m$: nominal productivity for the resource $m$. If there is only one type, the sub-index is avoided.

$\beta^{\omega}_{tm}$: effective worker $m \in \mathcal{W}$ productivity at time $t \in \mathcal{T}$ in scenario $\omega \in \Omega$ (kilograms per worker per period).

$\breve{\beta}_{t,m}$: actual deterministic productivity for the $m$ resource at moment $t$

$\xi \in \Xi : \{0,1\}$: the set of possible values that may take the uncertainty realization.

$\bar{\xi}_t$: expected realization in period $t$ (–) $t \in \mathcal{T}$.

$\dot{x}$: decorator for variables (i.e. $x$) that are decided before uncertainty of the period is realizated

$h^{\omega}_{jt}$: daily harvested quantity at $j \in \mathcal{J}$ block in period $t \in \mathcal{T}$ in scenario $\omega \in \Omega$, calculated as $\beta^{\omega}_t z^{\omega}_{jt}$ (kilograms/day).

$x^{\omega}_{tm}$: workers $m \in \mathcal{W}$ hired at time $t \in \mathcal{T}$ (workers) for path $\omega \in \Omega$ (when EV problem, $\Omega = \{1\}$, and belongs to the expected value of the uncertainty).

$y_{tm}$: workers $m \in \mathcal{W}$ laid off at time $t \in \mathcal{T}$ (workers) for path $\omega \in \Omega$ (when EV problem, $\Omega = \{1\}$, and belongs to the expected value of the uncertainty).

$\dot{z}_{jtm}$: workers $m \in \mathcal{W}$ allocated in block $j \in \mathcal{J}$ in period $t \in \mathcal{T}$ before uncertainty happens (workers) for path $\omega \in \Omega$ (when EV problem, $\Omega = \{1\}$, and belongs to the expected value of the uncertainty).

$z_{jtm}$: workers $m \in \mathcal{W}$ allocated in block $j \in \mathcal{J}$ in period $t \in \mathcal{T}$ after uncertainty is revealed (workers) for path $\omega \in \Omega$ (when EV problem, $\Omega = \{1\}$, and belongs to the expected value of the uncertainty).

$l^{\omega}_{mt}$: manpower or labor force $m \in \mathcal{W}$ at time $t \in \mathcal{T}$ (workers) for path $\omega \in \Omega$ (when EV problem, $\Omega = \{1\}$, and belongs to the expected value of the uncertainty).

$\mathcal{T}$: set of stages in the time horizon.

$j \in \mathcal{J}$: a specific block $j$ of the set of blocks of the vineyard.

$m \in \mathcal{W}$: a specific manpower resourcer $m$ of the complete set. If there is only one type, the subindex is avoided.

$r_{01}$: probability rain for two consecutives periods when first period is dry and the second is rainy

$r_{11}$: probability rain for two consecutives periods when both periods are rainy

$w^{\omega}$: conditional probablity for the specific scenario $\omega$

$a_j, b_j, c_j$: quality parameters for the cuadratic equation that represents the quality of the grape in the block $j$

$u_j$: fractional quality loss per rainy period for the grape in block $j$

$B_j$: price of the grape in lot $j$ (\$/kilograms).

$C_{E,m}$: cost of hiring (\$/worker).

$C_{F,m}$: cost to lay off (\$/worker).

$C_{P,m}$: cost of keeping labor idle between periods (\$/worker per period).

$C_{H.m}$: cost of harvesting (\$/kilograms).

$C_{Z,m}$: cost of assignment before uncertainty is revealed by worker $m \in \mathcal{W}$ (\$/kilograms).

$C_{\dot{Z},m}$: cost of assignment after uncertainty is revealed by worker $m \in \mathcal{W}$ (\$/kilograms).

$K$: maximum daily reception capacity of the winery (kilograms/day).

$S_j$ $j \in \mathcal{J}$: initial amount of grapes in lot $j$ (kilograms).

$Q^{\omega}_{jt}$: daily quality of the wine grape at $j \in \mathcal{J}$ block in period $t \in \mathcal{T}$ in scenario $\omega \in \Omega$ (–).

$\bar{Q}_{jt}$: average quality for that block $j \in \mathcal{J}$, $t \in \mathcal{T}$.

$\breve{Q}_{jt}$: actual deterministic quality for the block $j$ at moment $t$

$\mathbb{U}_{\mathcal{M}}$: expected value of the solution of a model, i.e., stochastic one.

$\mathcal{M}$: represents any model ant it is useful to write general expressions

$\xi^{\omega}_t$: the value of the uncertainty realization at moment $t$ for the specific scenario $\omega$

$t \in \mathcal{T}$: specific period time in the time span

$\boldsymbol{d}(i, j)$: the value of the nested distance algorithm where $i - th$ is the id number of the first tree to compare, and $j - th$ the id number of the second. It could be described shortly as $\boldsymbol{d}$

$\boldsymbol{d}(\tau)$: having fixed the two trees to be compared, it means the nested distance at any moment between them at time $t = \tau$, with $t \leq \mathcal{T}$

Note 1: when there is a unique type of resource available, the sub index $m$ is avoided.

Note 2: when there is a unique scenario (i.e., deterministic model), the supra index $\omega$ is avoided.

### 5.1. Rolling Horizon Approach

A multi-period decision model that considers the uncertainty in an explicit way, which may be represented by a tree and consequently, in the MS model. We denoted the complete and original tree as $\mathbb{T}$, with $\mathcal{T}$ time span size. The *rolling horizon approach*, RHA, is an algorithm that replaces the tree with a single scenario that is based on the expected probability of uncertainty. The optimization model is solved as a deterministic one using the artificial scenario, and the set decisions implemented. The optimization horizon for the RHA is smaller than the original; after some periods of decisions for policy execution, the system state was updated and the RHA optimization process repeated. We will go into further details on this approach, because it will make it easier to understand the rolling tree approach later.

The RHA considered that the original horizon of interest, $\mathcal{T}$, was approachable from successive optimizations where information on the state of the system is updated and the stochasticity is simplified. Considering the original tree, $\mathbb{T}$, with an uncertainty set of event $\Xi$ and probability distribution, $\mathcal{P}_t(\Xi)$ where $t \in \mathcal{T}$, the original tree could be simplified by replacing the probability distribution with an expected value, $\bar{\mathcal{P}}_t(\Xi)$, or briefly, $\bar{\mathcal{P}}$. The tree is reduced to a single sequence of events that are based on the expected value of the uncertainty; and the reduced tree with a single scenario is denoted by $\bar{\mathbb{T}}$. This simplification implies a loss of information that impacts on the final results. However, the bias depends on the uncertainty set, probability distribution and re-planning steps. We denoted each planning cycle as $k$, and $\mathcal{K}$ as the total number of cycles. The span for the planning time could reach the time span or be shorter than the original horizon. In our case, the planning time span was complete (finite), but decreased in length when we advanced through the planning steps. For a planning cycle $k$, its time span was $\mathcal{T}^k : (t_p^k : \mathcal{T})$, where $t_p^k$ is the initial period for the planning step $k$. The simplified tree for this initial planning cycle was denoted by $\bar{\mathbb{T}}^k$. It was optimized using initial information denoted by

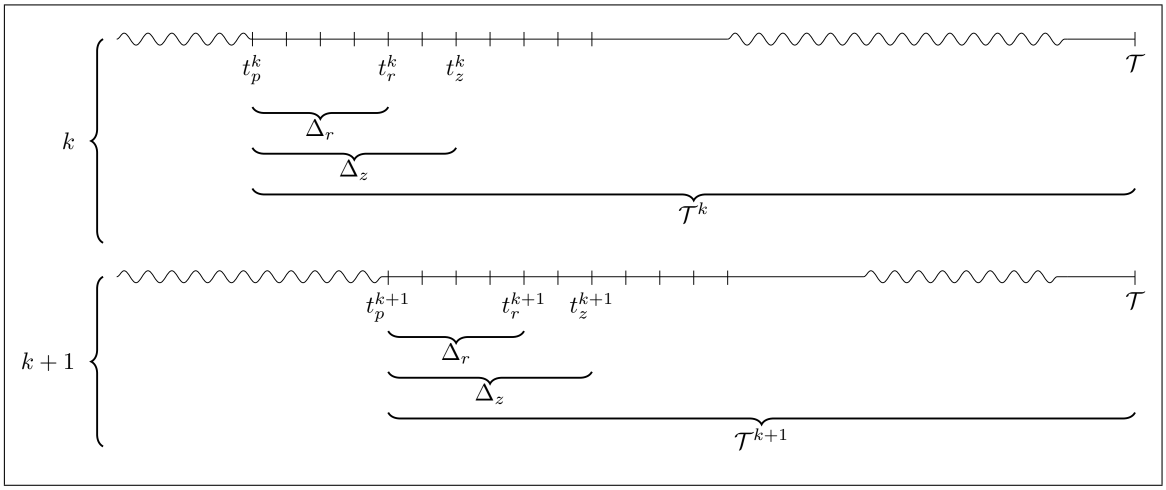



Figure 5.1. Schematic representation of two planning cycles for RHA

$\mathcal{I}^k$, to obtain a decision set $\mathcal{D}^k$. After the decision set had been applied by some periods in the system, the review time, $t_r^k$, is calculated, defined as $min\ (t_p^k + \Delta_r, \mathcal{T})$, where $\Delta_r$ is the review time span. At the review point, the decision-maker observes the state of the system and the information about it is updated; this information is the initial status of the system for the new planning cycle, $k+1$, $\mathcal{I}^{k+1}$. The next planning cycle begins at period $t_p^{k+1} = t_r^k + 1$. In some cases, the decision set $\mathcal{D}^k$ for the time span $\mathcal{T}^k$ will be applied even after the review time, $t_r^k$. The time span where decisions in the planning step $k$ will be applied is called the frozen time span, $\Delta_z$, and the last period of frozen decisions $t_z^k$ is defined as $min\ (t_p^k + \Delta_z, \mathcal{T})$. These decisions are rescued in every optimization, and are part of the decision policy for the next planning step. Considering two consecutive planning cycles, $k$ and $k+1$, the time parameters for the relationships are in eq. 5.1.

$$t_p^k \leq t_r^k \leq t_p^{k+1} \leq t_z^k \leq \mathcal{T} \tag{5.1}$$

To see more clearly the temporal parameters, we offer two planning cycles in Figure 5.1. To estimate the value of the RHA strategy, we run the complete planning process several times. We used simulations to represent the real events. The final profit from each complete planning process is the cumulative sum of the partial profits based on the implementation steps; for the $n \in \mathcal{N}$ running of the complete planning process, the partial profit from each planning step was $U_n^k$. To compare its performance with other models, the expected value is used, defined as $(\sum_{n\in\mathcal{N}}\sum_{k\in\mathcal{K}} U_n^k)/\mathcal{N}$, and written as $\mathbb{U}_{RH}$ as a general notation. In algorithm 1 is described the procedure for the RHA.

**Algorithm 1** Rolling Horizon Algorithm Expected Value

**Input:** Information: Instance, Initial conditions $\mathcal{I}$
**Input:** Model Information: $\mathbb{T}$, $\Xi$ , $\mathbb{P}$
**Input:** Structure: $\mathcal{T}$, $\Delta_z$, $\Delta_r$, $\mathcal{K}$
$n = 0$
**while** $n \leq \mathcal{N}$ **do**
  $U_n = 0$
  **for** $k = 1..\mathcal{K}$ **do**
    Determine $t_p^k$, $t_r^k$, $t_z^k$, $\mathcal{T}^k$ {Time Parameters}
    Determine $\bar{\mathbb{P}}^k \leftarrow \mathbb{P}^k$, $\bar{\mathbb{T}}^k \leftarrow \mathbb{T}^k$
    Determine $\mathcal{D}^k \leftarrow opt\ \mathbb{E}\ (\bar{\mathbb{T}}^k, \mathcal{D}^{k-1})$ {Optimization Routine}
    **Implementing:** {Through Simulation}
    **for** $t = t_p^k, ..., t_r^k$ **do**
      Uncertainty Realization $\hat{\xi}_t$
      State Update $\mathcal{I}_t^k \leftarrow \hat{\xi}_t, \mathcal{D}^k, \mathcal{F}, \mathcal{I}_{t-1}^k$ {Information update}
    **end for**
    Update $U_n + = U_n^k$
  **end for**
  Save final value $U_n$
  $n + = 1$
**end while**
$\mathbb{U}_{RH} = (\sum_{n\in\mathcal{N}} U_n)/\mathcal{N}$ {Expected Value Estimation }

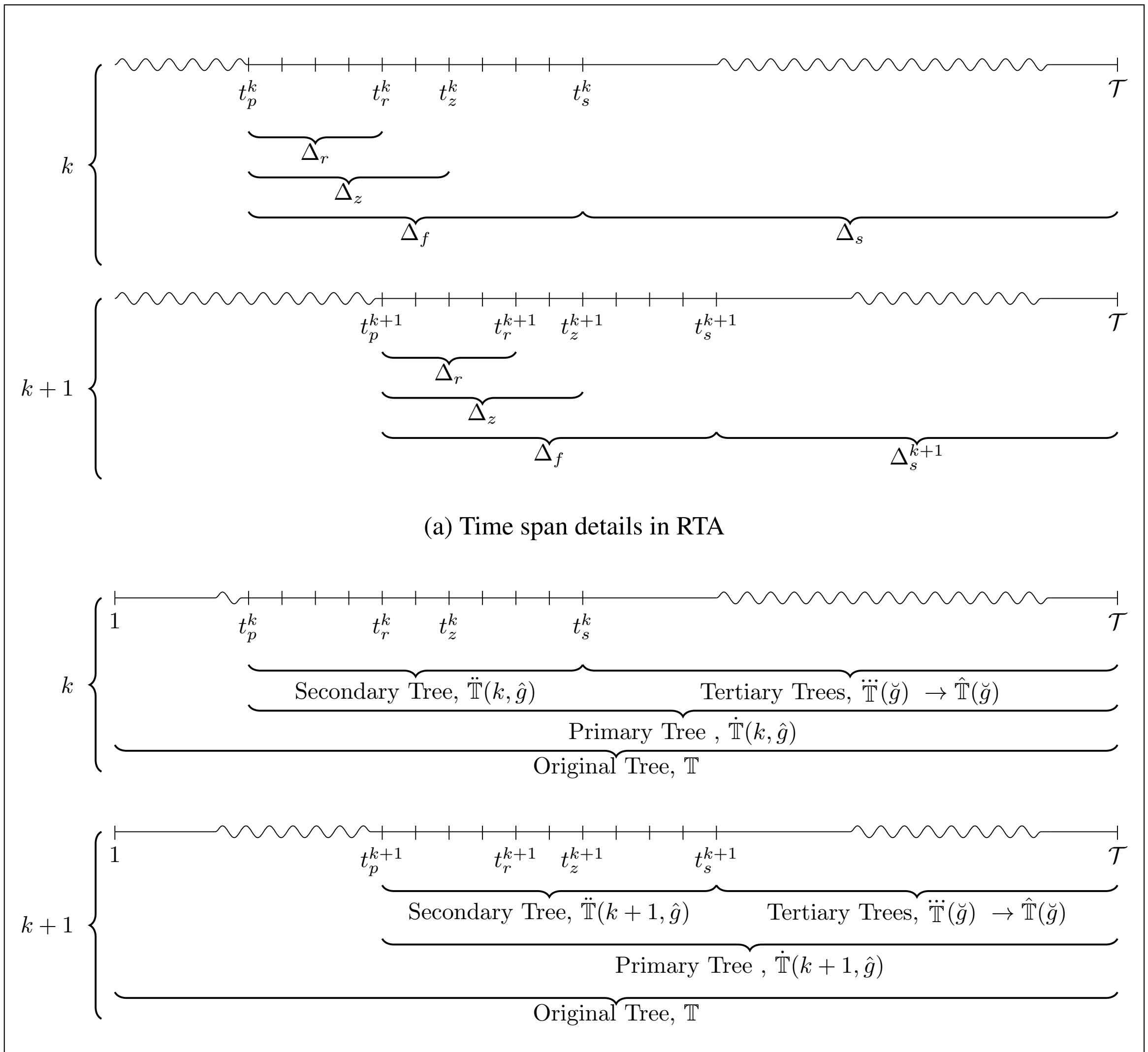
### 5.2. Rolling Tree Approach

We introduced the RHA in the last section; however, we now present the RTA based on RHA.

We denoted the *very original tree* as $\mathbb{T}$, with $\mathcal{T}$ time span size. The Rolling Tree algorithm, RTA, considers that the original horizon of interest, $\mathcal{T}$, is approachable from successive optimizations using a short term multistage model and a simplified view of the long term future. The RTA has similar principles to RHA, but is more complex especially because the information of the original tree is preserved to a larger degree. Each of the successive planning steps updated RTA information on the real state of the system. The progress of the system is obtained using simulations, giving the opportunity to visit different paths of realizations. The periods for the planning tasks are known as priori because all the time parameters are shared as information at the beginning.

Considering that our tree was now exhaustive, all the possible states and paths were perfectly listed. In each period there was a list of nodes where the information was kept. The list of nodes that belong to the planning periods are denoted by $\hat{g} \in \hat{\mathcal{G}}$, and the list of initial periods for planning, $\hat{t} \in \hat{\mathcal{T}}$.

In each of the planning steps, a new tree, shorter than the original is generated. These trees are called *primary trees* and represent a fan or sub-tree of the original one, retaining the original granularity and belonging to the planning step $k$ with an initial node $\hat{g}$, $\dot{\mathbb{T}}(k, \hat{g})$. The initial node is important because it contains the initial information, $\mathcal{I}(k, \hat{g})$. The initial period for the planning cycle $k$ is denoted by $t_p^k$. The primary tree extension is $[t_p^k, \mathcal{T}]$ and the original tree extension is $[1, \mathcal{T}]$, where $1 \leq t_p^k$. The first periods of $\dot{\mathbb{T}}(k, \hat{g})$ will be preserved as a tree, retaining the original bushiness, but the lasts are simplified. The part of the *primary tree* that keeps the original bushiness is called *secondary tree*, and is denoted by $\ddot{\mathbb{T}}(k, \hat{g})$. For the last period of $\ddot{\mathbb{T}}(k, \hat{g})$, there are final nodes that belong to the different leaves. Each of them is originally the birth of a new tree; and these trees will be simplified in our model, collapsing the granularity in a specific way. The list of the nodes that belong to the last period of the *secondary tree* are represented by $\breve{g} \in \mathcal{G}(k, \hat{g}, t_r^k)$. Each of the simplified trees is called *tertiary tree*, $\dddot{\mathbb{T}}(\breve{g})$; in this case the planning cycle is not relevant. The number of simplifications in each optimization process is $card(\mathcal{G}(k, \hat{g}, t_r^k))$. The simplified form of $\dddot{\mathbb{T}}(\breve{g})$ is written as $\hat{\mathbb{T}}(\breve{g})$. The first period of the simplified tree is $t_s^k$, so the two time spans are $[t_p^k, t_s^k)$ and $[t_s^k, \mathcal{T}]$. The simplified part of the tertiary tree could be done using different methods; we opted for an *expected value problem*, EV, and a *worst case problem*, WC. The main difference in the time parameters between RTA and RHA is that the planning horizon is divided in two parts for the RTA; the first part will be optimized as a fan or tree, and the remaining is simplified and solved as if the information was perfect. The time span to optimize as a fan is denoted by $\Delta_f$, such that the initial period in the simplified tree is $t_s^k = min\,(t_p^k + \Delta_f, \mathcal{T})$. The $min$ expression is linked to the last planning process, because the remaining time horizons could be shorter than the



(a) Time span details in RTA

(b) Trees details in RTA

Figure 5.2. Schematic representation of two planning cycles for RTA

original time spans. To clarify, Figure 5.2 offers two timelines that show the parts of the timeline that are encompassed by the different trees.

The RHA relationship presented in equation 5.1 is updated to equation 5.2. Figure 5.2 shows the timeline for RTA.

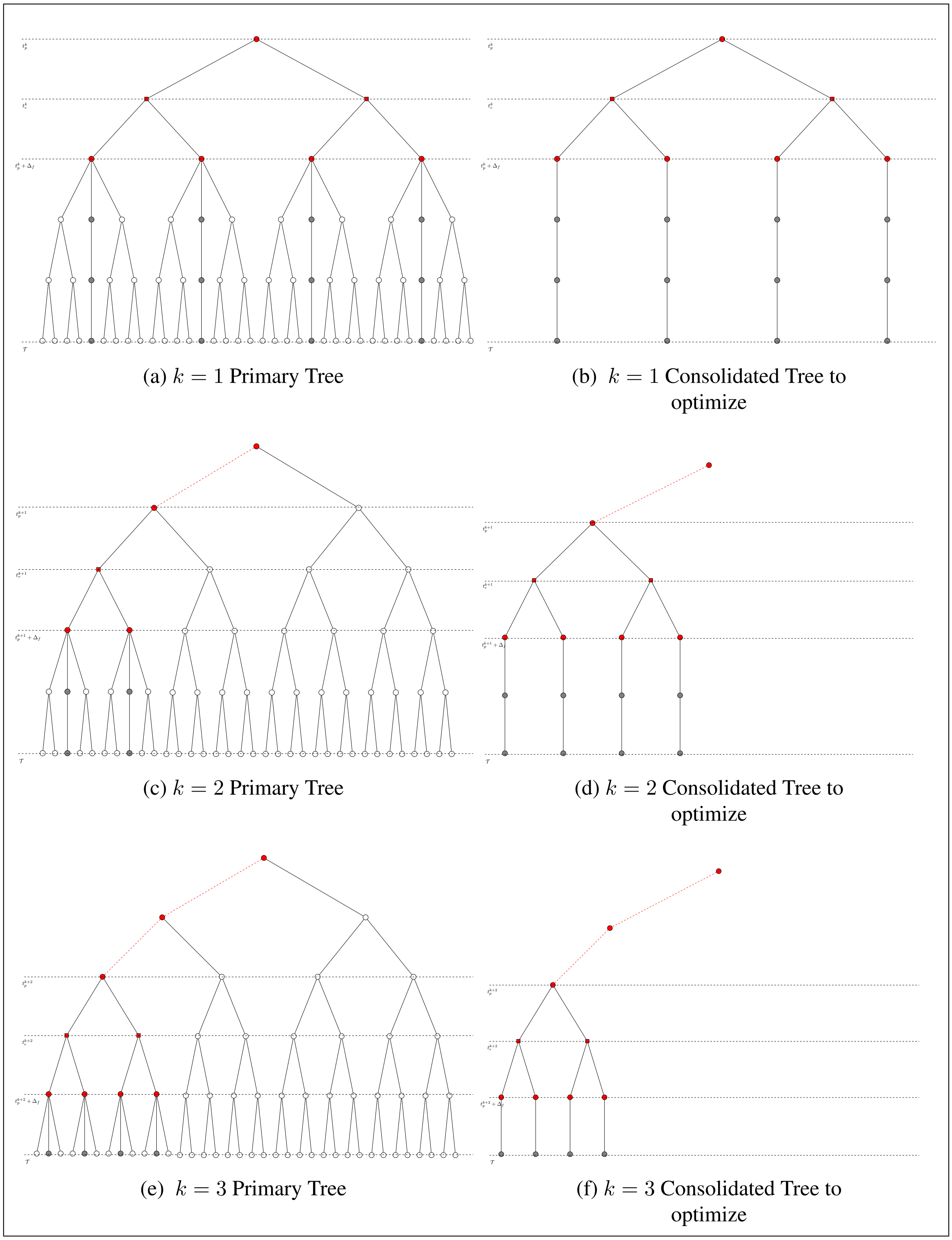
$$t_p^k \leq t_r^k \leq t_p^{k+1} \leq t_z^k \leq t_s^k \leq \mathcal{T} \tag{5.2}$$

The inequality $t_s^k \geq t_p^{k+1}$ means that the next planning cycle, $k+1$, will always be done on the fan part of the previous planning cycle, $k$. This is necessary because the nodes of the simplified part are not necessarily real. When we simplified the tertiary trees using the expected value, the optimized scenario probably does not exist; if the simplification criterion is WC, the scenario exists, even with low probability. The $t_r^k \leq t_p^{k+1} \leq t_z^k$ inequalities indicate that in the planning step $k+1$, previously-made decisions could exist; and in the $k$ step, decisions can be executed. This group of decisions will be conserved in the new planning step because the cost of changing them is prohibitive and is not instantaneous (RHA suffers the same).

In real life, each planning step is followed by an implementation time span and then a new planning step begins. In computational experiments, the realization is simulated. To clarify, in Figure 5.3, there is schematic example of how the rolling tree and its different elements roll over the tree according to the planning steps.

In Figure 5.3 (a), the primary tree for $k = 1$ is shown, maintaining the available total granularity. At this point, the first scenario simplification happens at period three. To represent them, we opted for the artificial scenario which had birth at the gray termination nodes of the secondary tree without branches. There are so many artificial scenarios as termination nodes are varied. In chart (b), the bushiness was erased to show the consolidated tree for the optimization of this planning step. Chart (c) shows a hypothetical realization of the uncertainty that was conducted at a specific node in period two, the moment of review, whose path is red and in a dashed line. The new primary tree for $k = 2$ is shorter than the original tree. A part of the original tree is not available to be visited, because the first realization leads to a path that is not linked with that side. It's important to mention that this is a representation following the tree graph; and it does not mean that the scenarios on the decisions could be overlapped, as well as the information at different paths. We prioritized the structure guide instead of the values, because the tour through the tree is based on the structure, and the rolling tree algorithm follows the same idea. The hidden side of the tree appears in a very light gray tone, and the available side in darker tones. This cycle continues until the final period is reached.

For computational implementation, the complete algorithm is shown in algorithm 2.



(a) $k = 1$ Primary Tree

(b) $k = 1$ Consolidated Tree to optimize

(c) $k = 2$ Primary Tree

(d) $k = 2$ Consolidated Tree to optimize

(e) $k = 3$ Primary Tree

(f) $k = 3$ Consolidated Tree to optimize

Figure 5.3. Schematic representation of planning steps for a binomial tree, with time parameter $(t_p^k, t_r^k, \Delta_f) = (1, 2, 3)$, using RTA

**Algorithm 2** Rolling Tree Algorithm

**Input:** Information: Initial conditions $\mathcal{I}$
**Input:** Model Information: $\mathbb{T}$, Simplification Strategy (EV or WC)
**Input:** Temporal structure: $\mathcal{T}$, $\Delta_z$, $\Delta_r$, $\Delta_f$, $\mathcal{K}$
**for** $k = 1 \ldots \mathcal{K}$ **do**
  Determine $t_p^k$, $t_r^k$, $t_s^k$, $t_z^k$ {Critical Periods}
  Load $\hat{g} \in \hat{\mathcal{G}}$ {Planning Node}
  $\dot{\mathbb{T}}(k,\hat{g}) \leftarrow \mathbb{T}$ {Primary Tree generation}
  $\ddot{\mathbb{T}}(k,\hat{g}) \leftarrow \dot{\mathbb{T}}(k,\hat{g})$ {Secondary Tree generation}
  Load $\breve{g} \in \mathcal{G}(k,\hat{g},t_r^k)$ {Terminal nodes list generation}
  **for all** $\breve{g} \in \mathcal{G}(k,\hat{g},t_r^k)$ **do**
    $\hat{\mathbb{T}}(\breve{g}) \leftarrow \dddot{\mathbb{T}}(\breve{g})$ {Simplified tertiary tree generation}
  **end for**
  $\hat{\dot{\mathbb{T}}}(k,\hat{g}) \leftarrow (\ddot{\mathbb{T}}(k,\hat{g}), \hat{\mathbb{T}}(\breve{g}),\ \forall \breve{g} \in \mathcal{G}(k,\hat{g},t_r^k))$ {Primary tree consolidation}
  $\mathcal{D}^k \leftarrow max\ \mathbb{E}\ (\hat{\dot{\mathbb{T}}}(k,\hat{g}), \mathcal{D}^{k-1})$ {Optimization Step}
  **Implementing:**
  $g_o = \hat{g}$ {Initial node for implementing step}
  **for** $t = t_p^k, ..., t_r^k$ **do**
    Uncertainty Realization $\hat{\xi}_t$
    Destiny node identification $g_d \in \mathcal{G}_t$
    Update of node information $\mathcal{I}(k, g_{d,t}) \leftarrow \hat{\xi}_t, \mathcal{D}^k, \mathcal{F}, \mathcal{I}(k, g_0)$
    Node id update $g_o \leftarrow g_d$
  **end for**
**end for**

### 5.2.1. Models Formulation

The RTA requires two considerations:

- On the simplifying procedure: EV simplification is done considering the way it was presented in Chapter 3. Basically, the tertiary trees are reduced using the expected value of the uncertain events to build a single scenario. This procedure is repeated for each terminal node of the secondary tree. For WC, the simplification is done making rain probability equal to one. In both cases, the secondary tree keeps the original probabilities and structure of the problem.
- About NACs: NACs are active in the secondary tree, but not in tertiary trees that are simplified. The reduced-tertiary-trees behave like independent path, so the decisions for similar periods are not necessary equals among reduced trees.

Next, we present the general optimization model for the RTA planning step. The model is very similar to the observed in Model 1, where the costs of recourse actions are negligible. In order to avoid confusing notation, part of that has been removed. Please read the model considering that it is for a planning step $k$ such as $1 < k < \mathcal{K}$; additionally, the structure of the primary-tree is $\dot{\mathbb{T}}(k, \hat{g})$.

$$
\begin{array}{lll}
\max & \sum_{\omega \in \Omega} w^{\omega} \sum_{j \in \mathcal{J}} \sum_{t \in \mathcal{T}} \{(B_j Q_{jt}^{\omega} - C_H) h_{jt}^{\omega} - C_E x_t^{\omega} - C_F y_t^{\omega} - C_P m_t^{\omega}\} & \\
s.t. & & \\
& m_t^{\omega} = m_{t-1}^{\omega} + x_t^{\omega} - y_t^{\omega} \quad \forall t \in \mathcal{T}^k,\, \omega \in \Omega & (rt1) \\
& \sum_{j \in \mathcal{J}} z_{jt}^{\omega} \leq m_t^{\omega} \quad \forall t \in \mathcal{T}^k,\, \omega \in \Omega & (rt2) \\
& \sum_{j \in \mathcal{J}} \beta_t^{\omega} z_{jt}^{\omega} \leq K \quad \forall j \in \mathcal{J}, t \in \mathcal{T}^k,\, \omega \in \Omega & (rt3) \\
& \beta_t^{\omega} z_{jt}^{\omega} \leq S_j - \sum_{t'=1}^{t-1} \beta_{t'}^{\omega} z_{jt'}^{\omega} \quad \forall j \in \mathcal{J}, t \in \mathcal{T}^k,\, \omega \in \Omega & (rt4) \\
& x_t^{\omega} = x_t^{\omega'} \quad \forall\, \omega', \omega \, \in \Omega_g,\, \omega' \neq \omega, g \in \mathcal{G}_{t-1}, t \in \mathcal{T}^k \,:\, 2 \leq t \leq min(\Delta_f, \mathcal{T} - kt_p^k) & (rt5a) \\
& x_1^{\omega} = x_1^{\omega'} \quad \forall\, \omega', \omega \, \in \Omega,\, \omega' \neq \omega & (rt5b) \\
& y_t^{\omega} = y_t^{\omega'} \quad \forall\, \omega', \omega \, \in \Omega_g,\, \omega' \neq \omega, g \in \mathcal{G}_{t-1}, t \in \mathcal{T}^k \,:\, 2 \leq t \leq min(\Delta_f, \mathcal{T} - kt_p^k) & (rt6a) \\
& y_1^{\omega} = y_1^{\omega'} \quad \forall\, \omega', \omega \, \in \Omega,\, \omega' \neq \omega & (rt6b) \\
& \dot{z}_{jt}^{\omega} = \dot{z}_{jt}^{\omega'} \quad \forall\, \omega', \omega \, \in \Omega_g,\, \omega' \neq \omega, g \in \mathcal{G}_t, j \in J, t \in \mathcal{T}^k \,:\, t \leq min(\Delta_f, \mathcal{T} - kt_p^k) & (rt7) \\
& x_t^{\omega}, y_t^{\omega} \geq 0, \in \mathbb{Z}_+ \quad \forall t \in \mathcal{T}^k,\, \omega \in \Omega & (rt8) \\
& \dot{z}_{jt}^{\omega} \geq 0, \in \mathbb{Z}_+ \quad \forall t \in \mathcal{T}^k,\, j \in J\,, \omega \in \Omega & (rt9)
\end{array}
$$

Model 6. Rolling tree optimization model for cycle $k$

### 5.2.2. Model Comparison Metrics

To measure the performance of RTA, we use three main indexes:

- Profit expected value of the solutions ($\mathbb{U}$)
- Decisions sets nested distances
- Computational times ($\theta$)

The expected value of the profit for MS and WS models is the objective function value. To refer to the expected value of the solution in RTA, we use the notation $\mathbb{U}_{RE}$ and $\mathbb{U}_{RW}$. The expected value is built using an algorithm that goes around all the planning nodes according to time structure, solving all the possible paths to the final period. The algorithm 3 shows the procedure that measure the RTA performance in order to be comparable to MS information.

**Algorithm 3** Algorithm to estimate the performance of the RTA in terms of expected value of the profit and rescue information

**Input:** Information: Initial conditions
**Input:** Model Information: $\mathbb{T}$, Simplification Strategy (EV or WC)
**Input:** Temporal structure: $\mathcal{T}$, $\Delta_z$, $\Delta_r$, $\Delta_f$, $\mathcal{K}$
**for** $k = 1..|\mathcal{K}|$ **do**
  Determine $t_p^k$, $t_r^k$, $t_s^k$, $t_z^k$ { Critical Periods }
  **for all** $\hat{g} \in \hat{\mathcal{G}}$ **do**
    $\dot{\mathbb{T}}(k, \hat{g}) \leftarrow \mathbb{T}$ {Primary Tree generation}
    $\ddot{\mathbb{T}}(k, \hat{g}) \leftarrow \dot{\mathbb{T}}(k, \hat{g})$ {Secondary Tree generation}
    Load $\breve{g} \in \mathcal{G}(k, \hat{g}, t_r^k)$ {Terminal nodes list generation}
    **for all** $\breve{g} \in \mathcal{G}(k, \hat{g}, t_r^k)$ **do**
      $\hat{\mathbb{T}}(\breve{g}) \leftarrow \dddot{\mathbb{T}}(\breve{g})$ {Simplified tertiary tree generation}
    **end for**
    $\hat{\dot{\mathbb{T}}}(k, \hat{g}) \leftarrow (\ddot{\mathbb{T}}(k, \hat{g}), \hat{\mathbb{T}}(\breve{g}), \; \forall \breve{g} \in \mathcal{G}(k, \hat{g}, t_r^k))$ {Primary tree consolidation}
    $\mathcal{D}^k \leftarrow max \; \mathbb{E} \; (\hat{\dot{\mathbb{T}}}(k, \hat{g}), \mathcal{D}^{k-1})$ {Optimization Step}
    **Implementing:**
    $g_o = \hat{g}$ {Initial node for implementing step}
    **for** $t = t_p^k, ..., t_r^k$ **do**
      Uncertainty Realization $\hat{\xi}_t$
      Destiny node identification $g_d \in \mathcal{G}_t$
      Update of node information $\mathcal{I}(k, g_{d,t})\,, U_{g_d} \leftarrow \hat{\xi}_t, \mathcal{D}^k, \mathcal{F}, \mathcal{I}(k, g_0)$
      $U_t^\omega \leftarrow U_{g_d}$ {Update information in $\mathbb{T}$}
      Node id update $g_o \leftarrow g_d$
    **end for**
  **end for**
**end for**
$\mathbb{U}_{RTA} = \sum_{\Omega, \hat{\mathcal{T}}} w^\omega U_t^\omega$ {Profit Expected Value}

The algorithm is called EVR, *expected value of the solution of RTA*. Used for RTE, we call the algorithm EVRE, and EVRW for RTW. The algorithm works as follows. Each node of the tree suffers several information refreshes as planning steps go ahead. Once the node is reached by the implementation procedure, the information is fixed, because the node will not be used in next planning steps. The EVRT solves the RTA for all the feasible planning nodes, $\hat{g} \in \hat{\mathcal{G}}$, and the information is displayed in a similar structure to the original tree. The outcome in the EVR is comparable to the monolithic outcome of the MS optimization model.

To value how different are the policy sets, we use the nested distance concept, introduced in Chapter 4. The information is rescued in the algorithm 3, when $\mathcal{I}(k, g_{d,t})$ is updated.

Finally, we hypothesized that for certain conditions RTA keeps an acceptable performance about the expected profit but with smaller computational time, represented by $\theta$. $\theta$(MS) is obtained from the solver optimization report. In that case, a complete tree is solved, so the computational effort should be greater than in the RTA, where the optimization stages are smaller about bushiness. Even when RTA needs multiple planning cycles to complete the whole process, the optimization time is individual for each of the cycle. We compare in this work, the optimization time for the first cycle that presents the same $\mathcal{T}$ of the MS model.

## 5.3. Results

### 5.3.1. Computational Experiments

In order to test hypothesizes, we run different numerical experiments. The cost parameters are similar to the describe in Chapter 3. The instance parameters are shown in Table 5.1. We introduce three-time structures, TS that provide different dynamic to the RTA; this is a short way of naming the temporal parameters as a unique concept. The relationship among the time spans in time structure is: $\mathcal{T} \geq \Delta_f \geq \Delta_z \geq \Delta_r$

It's important to highlight that the size of the fan is 80% of the original for TS1, 50% in TS2, and 20% in TS3. It means that computational effort should decrease according fan size decreases, but as it is mentioned in Rardin (2016), the final time depends on the instance and the structure of the problem.

Table 5.1. Parameters for the different experiments

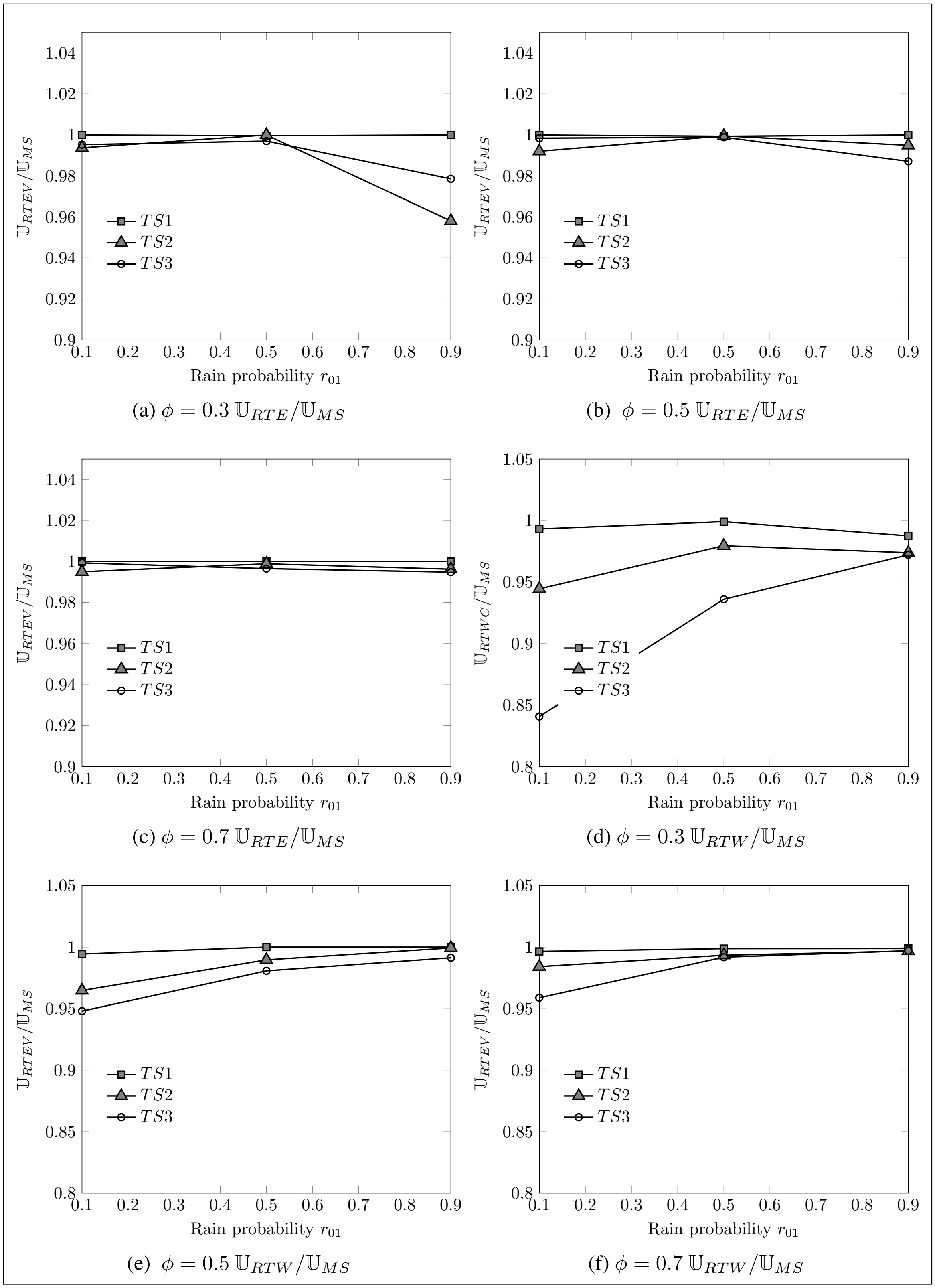
| Feature | Notation | Values | Units |
|---|---|---|---|
| Workers Ability | $\phi$ | 0.3, 0.5, 0.7 | – |
| Rain probability | $r_{01}$ | 0.1, 0.5, 0.9 | – |
| Rain factor between consecutive periods | $\tau_{11}$ | 0.7, 1.0, 1.3 | – |
| Rain quality penalty | $\gamma_j$ | 2% | % |
| Ripening Patterns | | SOD, MSH, HSM | – |
| Time Structure | TS1 | $(\mathcal{T}, \Delta_f, \Delta_z, \Delta_r) : (10, 8, 4, 2)$ | days |
| | TS2 | $(\mathcal{T}, \Delta_f, \Delta_z, \Delta_r) : (10, 4, 2, 1)$ | days |
| | TS3 | $(\mathcal{T}, \Delta_f, \Delta_z, \Delta_r) : (10, 2, 1, 1)$ | days |

The models were implemented using Python, written for PYOMO Python (Hart et al., 2017) and the optimization engine used was the GUROBI v. 8.1.0. The solution time was not limited, but the optimality gap was 1% and the integrality parameter was the solver default. We used a laptop computer with an Intel Core processor i7-6700HQ CPU 2.60 GHz, with 32.0 GB of RAM memory running Windows 10. We tested a total of 243 instances per model, RTE, RTW, RH, WS, EV and MS. The total number of optimization running was 1458. In this chapter we report only the main results, especially the comparison among RTA variants and MS.

### 5.3.2. Economic Performance of RTE and RTW

In Figure 5.4, the relative expected profit value is presented for models RTE, RTW and MS. The figure has six charts. The chart $(a)$ shows the results when ability is low, $\phi = 0.3$; the second one, $(b)$, shows the intermediate ability results, $\phi = 0.5$, and the last, $(c)$, is referred to high levels of ability, $\phi = 0.7$. The same results are shown for RTW in charts $(d)$, $(e)$ and $(f)$, respectively.The x-axis in the three charts shows the rain probability $r_{01}$. All the cited charts are for RTE. The last three charts are referred to RTW. Finally, in each of the charts, three series are shown, TS1, TS2 and TS3.

Figure 5.4 $(a)$ shows the $\mathbb{U}_{RTE}/\mathbb{U}_{MS}$ value for $\phi = 0.3$ or poor skill. For TS1, the fan time span is similar to the original time horizon, and the expected performances between




(a) $\phi = 0.3\ \mathbb{U}_{RTE}/\mathbb{U}_{MS}$

(b) $\phi = 0.5\ \mathbb{U}_{RTE}/\mathbb{U}_{MS}$

(c) $\phi = 0.7\ \mathbb{U}_{RTE}/\mathbb{U}_{MS}$

(d) $\phi = 0.3\ \mathbb{U}_{RTW}/\mathbb{U}_{MS}$

(e) $\phi = 0.5\ \mathbb{U}_{RTW}/\mathbb{U}_{MS}$

(f) $\phi = 0.7\ \mathbb{U}_{RTW}/\mathbb{U}_{MS}$

Figure 5.4. Relative Expected Profit Value for three different time structure and resource flexibility - SOD pattern

the models do not differ mostly. TS2 offers a shorter fan, around 50% of the original time, and more planning steps. It introduces a small loss in the simplified model, but this is still less than 4% when rain probability is very high. When the number of planning steps increases, reaching the daily information update, TS3, $\mathbb{U}_{RTE}/\mathbb{U}_{MS}$ is close to 1. To improve abilities, one needs to consider an exhaustive tree or update the information frequently. The RTE with highly frequent updates, does not require granularity because the tomorrow's decisions, and for an expected future are made today. If rain probability is low in the future, it will not affect the performance of RTE. If rain event is highly probable, then the expected value considers mostly the negative occurrence and the decisions will be conservative, losing the opportunity to capture better alternative although they are few probable. While shorter the time span for revision, better reactions could have the system in extremely negative conditions if approximation is used; however, the implementation or frozen horizon should be short because the information update requires to be active in the decision model.

In Figure 5.4 charts $(b)$ and $(c)$, the same information is presented for medium and high ability, respectively. Regardless of the time structure, both graphs support the Chapter 3 conclusions; higher ability in the resource takes off value from the model contribution, making the last less critical. **At medium or major levels of ability, the performance of RTE seems to be comparable to MS.**

The last three charts in Figure 5.4 shows the information for RTW. The worst case considers that in the future will be always raining. This affects the quality and creates a context of risk about quality of the product. This standpoint punishes low rain probabilities in the simplified tertiary trees, but it works, intuitively, well when rain probability is really high. In the same way, for low rain probabilities, it is desirable to have less number of planning steps to avoid the negative bias of the simplification.

For the three levels of skills, when partial fan is similar to the original one (MS), series TS1, the relative gaps are minors, in terms of expected values.

As the partial fan becomes shorter, the gap with MS increases. The worst situation, with losses about 10% is when the rain probability is very low, and the fan is very short. Two situations impact on this gap: first, the granularity of the events is very poor, and the weight of the expected scenario is very high in each optimization step. Second, even when rain is very unlikely, the simplification considers the event occurrence as sure. This increases the costs, so the strategy is very different to the MS case. Furthermore, when rain event is highly probable, the gap decreases considerably. It gives an interesting clue: the rain event or uncertainty, plays a critical role in the relative value of RTW. In Figure 5.4 $(e)$ and $(f)$ the gap between models decreases because of the contribution of the resource flexibility.

### 5.3.3. Quality Value in RTA approach

The ripening pattern may change the decisions and the value of the system consequently. We presented formerly the SOD patterns results, now we show the results for the HSM pattern.

Figure 5.5 shows that the HSM pattern offers competence between RTE, RTW and MS approaches. The differences saw in SOD pattern for these models, do not appear here. In that case, was the high rain probability where differences make greater. Here the ripening pattern changes and then the value over time, specifically, is more concentrated at the beginning. With this scenario, even when granularity is low, this reaches the first period and then the value is mostly captured.

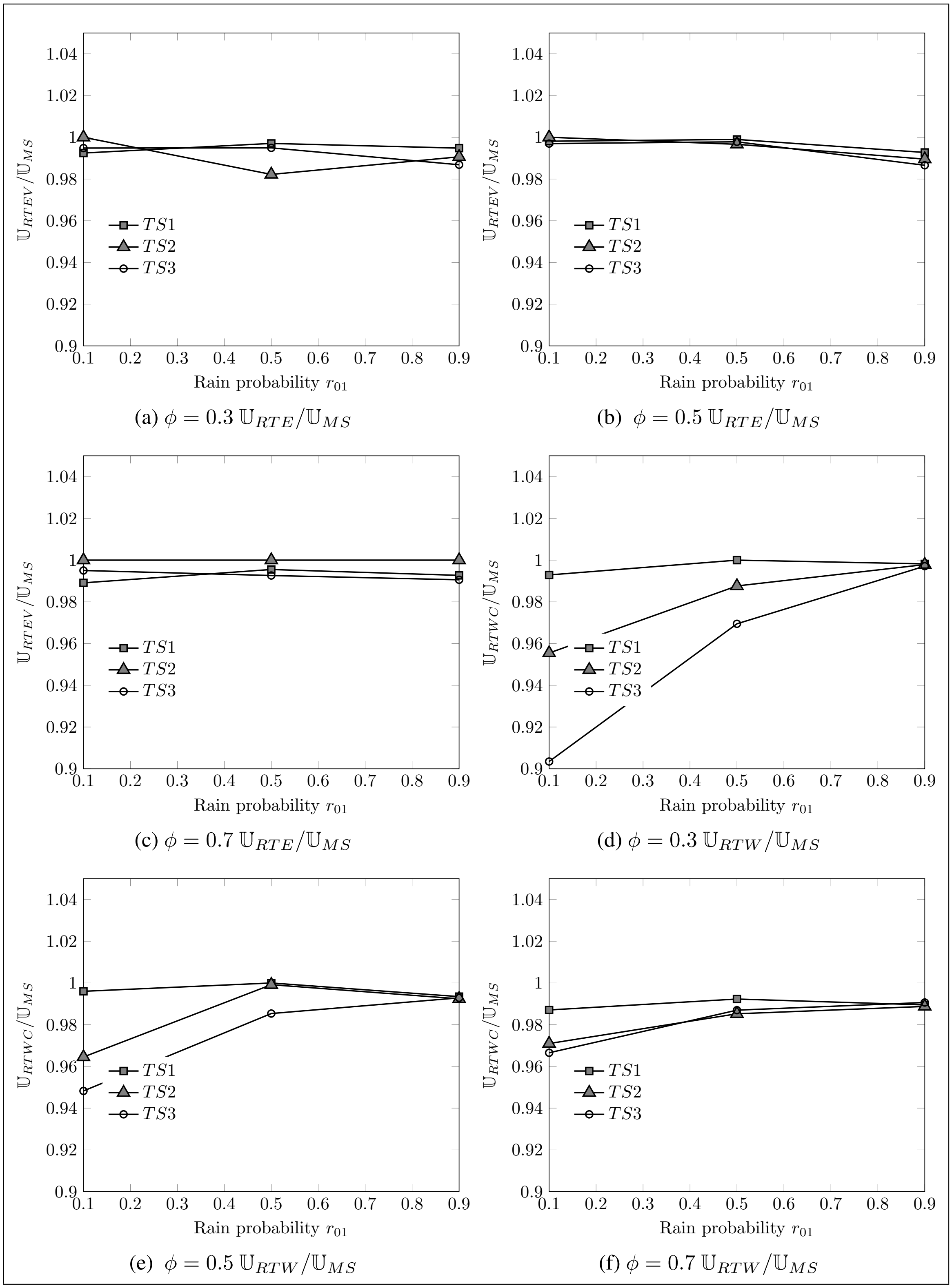
Figure 5.5 shows in $(d)$ a loss about 10% when rain probability is very low and the partial fan time is far way of the original one. The gap is greater than in the same case for SOD pattern; the reason is that HSM is a negative scenario for the ripening process as we indicated in Chapter 3, so the stress is really high, and the better performance of MS model is emphasized. When the resource ability increases, the results are better, but keeping a worse performance at low rain probabilities.

As gaps are very short in some cases, we present the values of the four previous graphs in Table 5.2.

### 5.3.4. Manpower Policy Nested Distance

In order to discuss the manpower decisions policies obtained by the models, we compared them using the nested distance.

In Figure 5.6 we observe the nested distance for the completes trees where the probability structure is similar to the original tree, and nodes information are the decisions made under




(a) $\phi = 0.3$ $\mathbb{U}_{RTE}/\mathbb{U}_{MS}$

(b) $\phi = 0.5$ $\mathbb{U}_{RTE}/\mathbb{U}_{MS}$

(c) $\phi = 0.7$ $\mathbb{U}_{RTE}/\mathbb{U}_{MS}$

(d) $\phi = 0.3$ $\mathbb{U}_{RTW}/\mathbb{U}_{MS}$

(e) $\phi = 0.5$ $\mathbb{U}_{RTW}/\mathbb{U}_{MS}$

(f) $\phi = 0.7$ $\mathbb{U}_{RTW}/\mathbb{U}_{MS}$

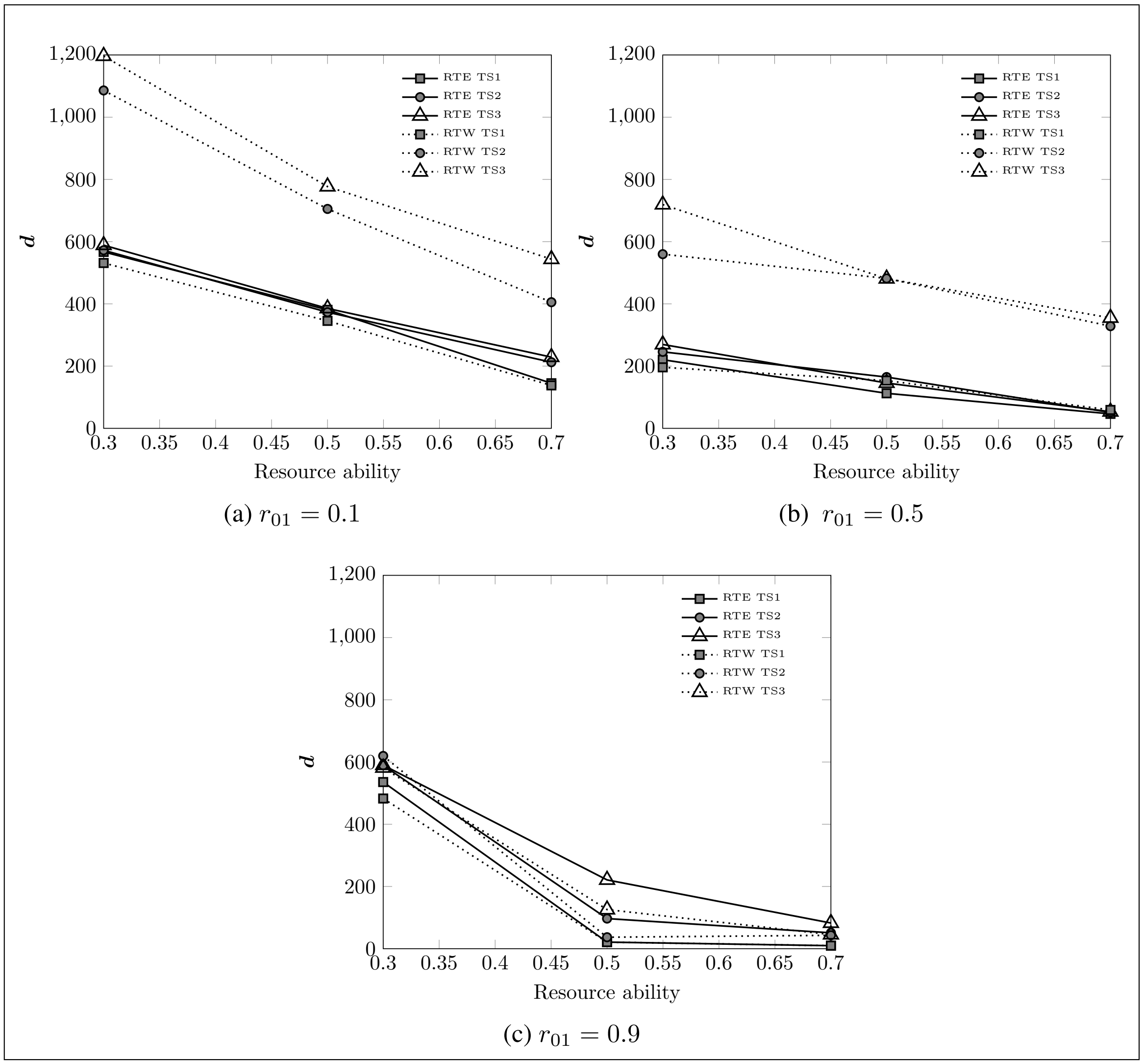
Figure 5.5. Relative Expected Profit Value for three different temporal structure and resources flexibility for HSM pattern

Table 5.2. Relative Expected Profit Value

| | | | $\mathbb{U}_{RTW}/\mathbb{U}_{MS}$ | | | $\mathbb{U}_{RTE}/\mathbb{U}_{MS}$ | | |
|---|---|---|---|---|---|---|---|---|
| | | $r_{01}$ | 0.1 | 0.5 | 0.9 | 0.1 | 0.5 | 0.9 |
| SOD | $\phi = 0.3$ | TS1 | 0.993 | 0.999 | 0.988 | 1 | 1 | 1 |
| | | TS2 | 0.944 | 0.979 | 0.974 | 0.994 | 1 | 0.958 |
| | | TS3 | 0.841 | 0.936 | 0.972 | 0.995 | 0.997 | 0.979 |
| | $\phi = 0.5$ | TS1 | 0.994 | 1 | 1 | 1 | 0.999 | 1 |
| | | TS2 | 0.965 | 0.99 | 0.999 | 0.992 | 1 | 0.995 |
| | | TS3 | 0.948 | 0.981 | 0.991 | 0.998 | 0.999 | 0.987 |
| | $\phi = 0.7$ | TS1 | 0.996 | 0.999 | 0.999 | 1 | 1 | 1 |
| | | TS2 | 0.984 | 0.993 | 0.997 | 0.995 | 0.999 | 0.996 |
| | | TS3 | 0.959 | 0.992 | 0.997 | 0.999 | 0.997 | 0.995 |
| HSM | $\phi = 0.3$ | TS1 | 0.993 | 1 | 0.998 | 0.992 | 0.997 | 0.995 |
| | | TS2 | 0.955 | 0.988 | 0.998 | 1 | 0.982 | 0.991 |
| | | TS3 | 0.903 | 0.969 | 0.997 | 0.995 | 0.995 | 0.987 |
| | $\phi = 0.5$ | TS1 | 0.996 | 1 | 0.993 | 0.998 | 0.999 | 0.993 |
| | | TS2 | 0.965 | 0.999 | 0.992 | 1 | 0.997 | 0.99 |
| | | TS3 | 0.948 | 0.985 | 0.993 | 0.997 | 0.998 | 0.987 |
| | $\phi = 0.7$ | TS1 | 0.987 | 0.992 | 0.989 | 0.989 | 0.996 | 0.993 |
| | | TS2 | 0.971 | 0.985 | 0.989 | 1 | 1 | 1 |
| | | TS3 | 0.967 | 0.987 | 0.991 | 0.995 | 0.993 | 0.991 |

each type of models and time structures. We use the MS model as reference to calculate the ***d***, so we will simplified the notation leaving the MS subscript, i.e., from ***d***(MS,RTW) to ***d***(RTW).

Chart $(a)$ shows different cases of ***d***(RTE) and ***d***(RTW). Both RTA simplifications lead to a minor distance to MS when resources ability increases. It is reasonable because high capabilities or abilities make decisions less dependent on uncertainty, so the exhaustive tree probably offers information that is not as critical. In the RTE case, the time structure seems not to have a special weight in the trend and magnitude of the distance. Values are concentrated with little dispersion at each level of resource flexibility.



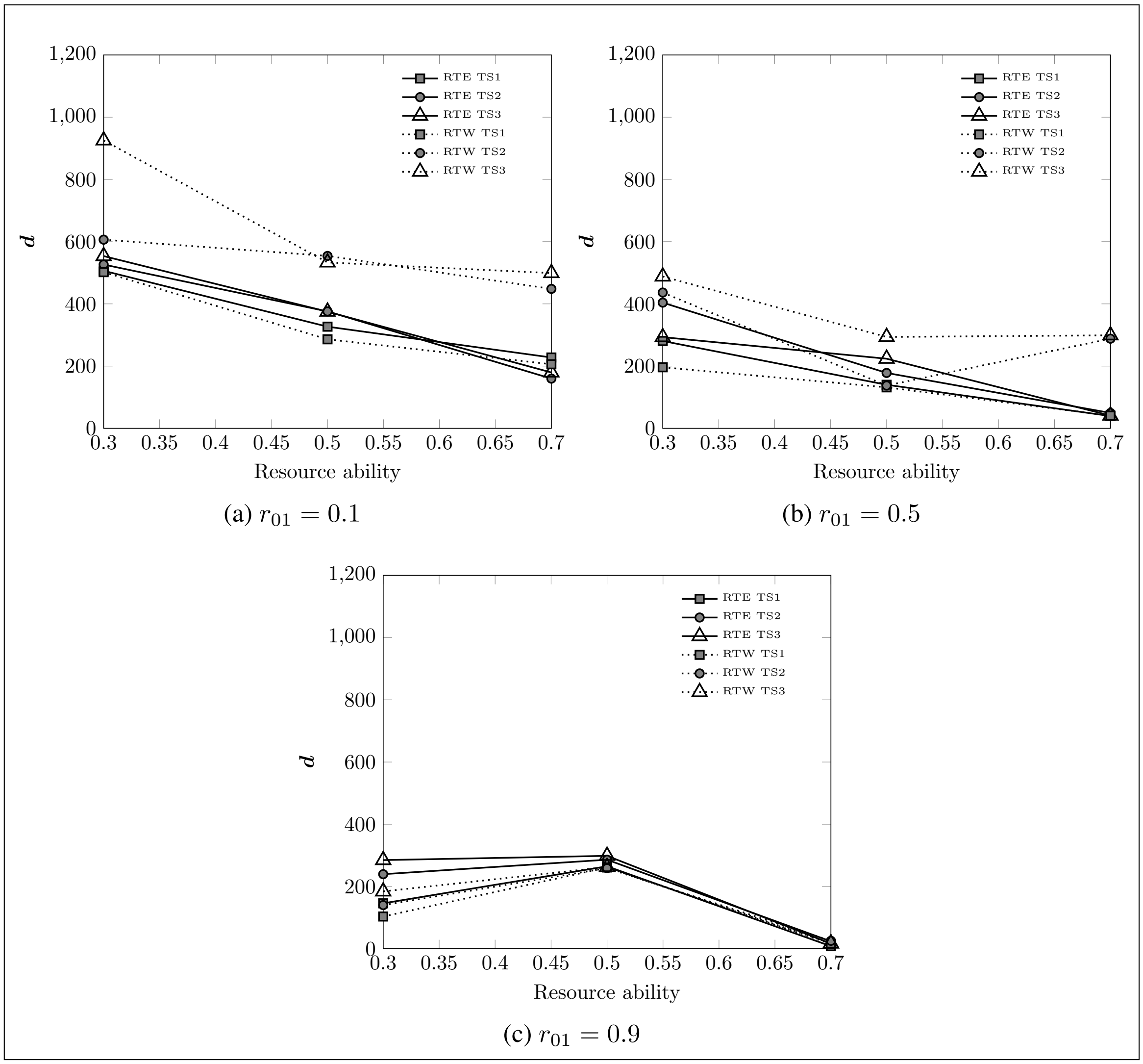

(a) $r_{01} = 0.1$ (b) $r_{01} = 0.5$

(c) $r_{01} = 0.9$

Figure 5.6. Final nested distance value (y-axis) for three different time structure (series) and resource's ability (x-axis). Both RTA are in these charts. The rain probability is low in chart $(a)$, medium in chart $(b)$ and high in chart $(c)$. The quality pattern is SOD

For TS1, the $\boldsymbol{d}$(RTW) is pretty similar to the $\boldsymbol{d}$(RTE). Both consider the same information for the fan part, and this is almost the same as MS original model. The gap between them is related to the probabilities considerations in the simplified part. When granularity of the secondary tree decreases, $\boldsymbol{d}$(RTW) increases.

In chart $(b)$, uncertainty is maximum. Both RTA distances decrease compared to $r_{01} = 0.1$ chart. In the case of RTE, the probability simplification leads to behaviors that are not so different to the MS approach. We know that in the case of EV models, the performance




(a) $r_{01} = 0.1$

(b) $r_{01} = 0.5$

(c) $r_{01} = 0.9$

Figure 5.7. Total nested distance value (y-axis) for three different time structure (series) and resource's ability (x-axis). Both RTA are in these charts. The rain probability is low in chart $(a)$, medium in chart $(b)$ and high in chart $(c)$. The quality pattern is HSM.

difference is largest when uncertainty also (Chapter 3); this effect is weakened by the update of information. The RTW simplification leads to worse trees when granularity of the secondary tree decreases. The ability impact is present in all the approximations.

The last chart, $(c)$, is for a highly rainy probability. In this situation, $\boldsymbol{d}$(RTW) and $\boldsymbol{d}$(RTE) behave very similar. Increasing granularity of secondary tree, diminishes the distance to MS, and higher abilities reduce also the differences between the sets of decisions.

In Figure 5.7 shows the same information that in Figure 5.6, but for HSM ripening pattern in this case. The ability impact trend follows the behavior that we described in Chapter 3 and also here. When $r_{01} = 0.1$, the nested distance values in RTE are concentrated in different levels depending on the ability of the resource. RTW is more variable, strongly affected by the granularity of the secondary tree and the bad approximation made for the rain event. When rain probability is strongly uncertain, the differences between the policy trees are fewer than in the SOD case. In this pattern, the highest value is in the first periods, so event with a minor granularity in the secondary tree, the value of the system could be captured. Finally, chart (c) shows the nested distance value for very probable rain environment. There is a difference among models when ability is low, something predictable.

### 5.3.5. Computational Times

To complete the performance analysis of RTA, we compare now the computational effort. MS has been said to be a huge consumer of resources when size increases. RTA should offer an advantage in that sense without great performance losses. Table 5.3 shows a summary of the quartiles of the recorded optimization times, where the last two rows offer two relative indicators comparing the RTA behavior and MS performance. All the information is expressed in milliseconds, excepts the relative performances rows.

Table 5.3. Computational times performances. Optimization time expressed in millisecond

| | $\theta$(RTE) | | | $\theta$(RTW) | | | $\theta$(MS) |
|---|---|---|---|---|---|---|---|
| Statistics | TS1 | TS2 | TS3 | TS1 | TS2 | TS3 | |
| Scenario Number | 256 | 16 | 4 | 256 | 16 | 4 | 1024 |
| Min | 28.25 | 14.04 | 10.25 | 268.94 | 45.79 | 42.26 | 1673.03 |
| Median | 33.28 | 46.89 | 19.62 | 336.67 | 52.01 | 106.19 | 2415.74 |
| Max | 144.73 | 184 | 145.63 | 419.92 | 72.29 | 500.6 | 47955.59 |
| Average | 54.58 | 52.91 | 34.6 | 337.9 | 139.14 | 53.11 | 6280.79 |
| RT average / MS average | 0.87% | 0.84% | 0.55% | 5.38% | 0.85% | 0.62% | |
| RT median / RT average | 1.38% | 1.94% | 0.81% | 13.94% | 2.15% | 4.40% | |

Both RTW and RTE show that the simplification expected speeds are always smaller than MS one's. The behavior is not homogeneous, but the magnitudes are very different to MS. Time structures, TS, should be a good estimator of the final speed. In the case of RTE, the TSs do not present important differences among configurations. In RTW, the differences are more important, and the consumed time is up five time the RTE computational effort. However, in both cases, there are less than 5% of the MS optimization time.

## 5.4. Last Stage Distributions

The final analysis is about the distributions of the outcome in RT and MS. As we noted before, the expected profit value is a performance index, but it does not give any details about the real distribution of the individual scenarios and probable results of the model. Individual scenarios performance and their probability of happening hide risks for the decision-maker. To observe this behavior, we opt by expected shortfall, ES, where poor skilled labor ($\phi = 0.3$) is confronted in the complete set of models for two conditions of rain probability. Resulting distribution are shown in Table 5.4. The results show that ES for RTA variants are worse than MS cases, but negative scenarios are not present in a statistical representative way.

Table 5.4. Expected Profit Value ES$_{10\%}$ [\$(000) and relative expected profit, ES$_{RTA}$/ES$_{MS}$

| | RTE | | | RTW | | | MS |
|---|---|---|---|---|---|---|---|
| ES | TS1 | TS2 | TS3 | TS1 | TS2 | TS3 | |
| $r_{01} = 0.1$ | 1375 | 1567 | 1599 | 1630 | 1614 | 1615 | 1698 |
| $r_{01} = 0.9$ | 211 | 218 | 228 | 139 | 208 | 139 | 233 |
| Relative $r_{01} = 0.1$ | 81.0% | 92.3% | 94.1% | 96.0% | 95.0% | 95.1% | |
| Relative $r_{01} = 0.9$ | 91% | 94% | 98.5% | 60% | 89% | 60% | |

## 5.5. Conclusions

In this chapter we introduce the rolling tree approach, RTA, hypothesizing that this approximation could offer a good approach to the final expected value of MS approach, with less computational effort. First, we describe the rolling horizon approach, RHA, giving a complete description of the elements of the algorithm and the way it works; immediately after we introduce the rolling tree algorithm. Then, the basis of the optimization model and metrics to be used in this chapter are shown. Subsequently, results are presented.

The main purpose of Chapter 5 was to discuss if a kind of rolling horizon approximation may be a better solution for those cases where EV fails in terms of profit performance. We chose the specific conditions where EV was not successful, providing big gaps between EV and MS.

In RTA, we see that for SOD pattern, when ability increases, the RTA-MS economic performance gap decreases. RTE behaves very close to the MS model, showing good results for different levels of rain probability. Granularity of the secondary tree and planning final number of cycles seem not to have a great effect in the final performance, losing few values compared to MS. In RTW, the losses are more important when rain probabilities are low, because of the poor approximation to those cases. When rain probability increases, the RTW-MS gap becomes smaller and even acceptable in some cases. The quality degradation plays a critical role; for HSM pattern, the value is concentrated at the first periods, so the granularity of secondary tree and the few feasible paths at the first periods, impacts positively on the capture of the higher prices, and finally, value.

In terms of computational times, RTA is considerably faster than MS model. The gain depends on the number of planning steps and their sizes. The rain probability also plays a role here, changing the shape of expected times.

The last comparison for the original test time structures and instances was how significant are the differences between the set of decisions made under different models. The results indicate that as the rain probability increases, at least for the examined conditions, the decisions trees trend to be closer, but not equals. It iss imperative to mention that probabilities are similar, so the differences are concentrated in the nodal decisions.

The last part of this chapter is about the distribution of the final results of the scenarios. RTA has a partial myopic vision about future, so some decisions may lead to negative scenarios that are hidden in the average performance analysis. We observed for a pair of examples that the exposure to negative results are nonexistent both in RTA and MS.

However, it is only an example of two rain probabilities instances, and more research should be developed to study more general the risk under RTA models.

To conclude, we consider that RTA has offered promising results, in accordance with our expectations, that encourage the continuity of studies in this research line. Key information about the conditions in which they develop their potential and a magnitude of the probable impacts that are very important for a manager has also been presented. A sensitivity analysis may add notable information for applications. However, we consider that this kind of fine-tuning should be done in the specific application because costs play an important role. We highlight the way of conducting the experiments, in order to understand better each of farmer's unique case. Finally, we believe that a *Decision Support System*, DSS, would be a very useful step to get close to the real application.

# Chapter 6

# Conclusions and Future Work

This chapter is a summary of the thesis. First, we describe what has been done. Then, the most relevant conclusions are exposed, and finally, the future research directions are presented.

### 6.1. Thesis overview

Agricultural planning has been recognized as especially complex given the number of sources of uncertainty that are present. The market structure, biology of the products, and the management of industrial and field operations present high stochasticity.

In this work we addressed the wine grape case, an especially important agricultural item in Chilean market. Wine grape is responsible in great part for having excellent qualified wines in the international context, so its quality should be the object of study under the right management. Expert judgment is critical, but the complexity of decisions consequences over time would be benefited with the help of mathematical supports.

Our research was focused on the multi-period models where uncertainty was critical. The base formulation was a multi-stage stochastic model, a contribution to new research literature where the variability consideration has not been addressed extensively. This model describes the farmer decisions in the harvest time, by means of a fundamental moment for the final quality. The farmer's dilemma can be summarized as: to anticipate the harvest destroys value because the optimum maturity (the maximum price) is still not reached, but this anticipation decreases the risks of negative impacts because of rain consequences, the uncertain event. Hence, anticipating or postponing needs an economic analysis but also an operational interpretation of the feasibility of the decisions to be implemented.

At the beginning of this thesis, we reviewed the literature to look for multi-stage stochastic programming applications in agriculture, and even more in fresh fruit cases (Chapter 2)

The first approach to the problem was to compare an expected value problem with multistage stochastic programming (Chapter 3. The EV considers the uncertainty in an expected way, creating a scenario used in deterministic optimization. While EV recognizes the uncertainty existence (deterministic does not), the posterior transformation implies loss of information, so it has been described as having poor performances against stochastic approaches. Multistage stochastic programming was our first stochastic approach. It describes as a binomial tree, the possible paths in an exhaustive way. Initially, two versions are presented: with recourse and without recourse actions. The difference between them is that not all the decisions could be made after the uncertainty was revealed (second epoch). Making decisions after uncertainty has been revealed to be flexibility source because it allows the farmer to react after information is updated. It also means added costs, because if flexibility was free it would be a commodity and not a competitive advantage. The EV model was compared to MS in terms of the expected profit with and without recourses. In order to understand how ripening patterns change the value of the system, we tried three different configurations with different optimum ripening moments.

Chapter 4 explores in detail the effect of the decision epochs and other configuration of resources. We formally introduced the flexibility concept defined as the ability to react to a variable context. Having flexibility implies paying an extra cost that will be recovered only if it could be used. We tested three different models to understand the epoch of decision-making; which could be before, after, and before with later corrections of the uncertainty realization. Thus, each of the allocation moments has different costs to value the extra information. Additionally, we tested the possibility of constituting teams with two types of resources, with different productivity levels and similar skills to work under the rain.

After several optimization runs, the computational effort was found to be an important factor to consider, because of the hardware requirement (see the exponential behavior with the size increment of the model, Chapter 3). This problem has been reported in the literature like the tractability problem and different methods, especially since the reduction tree has been used. In Chapter 5 we face this problem using the rolling horizon approach in a novel algorithm called rolling tree. The hypothesis is that the original complete tree could be split in minor trees that rolls over time and maintains a level of granularity for short-term decisions and simplified view for the rest of the original time span of optimization. Intuitively, if the number of leaves decreases then the optimization time goes in the same direction. The cost of this simplification is the loss in quality of the decisions, but it is not clear in which conditions this kind of approximation should be used with a controlled loss. To explore these balances, we designed an algorithm that gives form to the rolling tree approach, and then a way of measuring the expected profit to be compared with the MS results, our benchmark. We run different time structures that are critical to create the rolling approach, and we compared three performances, expected final value, difference among trees of decisions and computational effort in terms of the expected time. To simplify the tree in the final periods, we used the expected value (RTE) and the worst case used in other model, RTW.

Finally, we considered that the RT approach could hide negative scenarios under an acceptable expected value performance. We reviewed some of the final distributions and the existence of negative scenarios was discarded for at least four of our experiments. This does not mean that risk does not exist in RT approach, but it needs more development that exceeds this initial study.

## 6.2. Conclusions

The general conclusions of this thesis are listed below:

i The expected economic performance gap between EP and MS models, is a function of the uncertainty and the flexibility of the resources of the system to face the uncertain event. While higher skilled resources are used, the uncertainty impact diminishes and the decisions support model adds extra value to the system. EP is very useful in this situation due to the lower computational effort required to solve the problem in a correct way, giving a decision policy that is stable over time.

ii When the uncertainty was at a maximum (rain and no-rain are equally probable), for medium and low ability in resources, the MS approach creates considerable value and the gap with the EV problem increases. It could be explained assuming that the information for decision-making is not biased enough, hence, the application of the EV decision policies is not appropriated for any of the feasible paths.

iii The decisions set for the models differ in capacity to anticipate or postpone decisions. EV considers a perfectly known future, so the reaction to not describing events is null. In MS without recourse, MB, anticipated decisions are used in order to avoid risky future scenarios.

iv The MA model is not always the best to apply. The cost in some cases is not recover (in expected terms). For high or medium ability, MB could be a very good approximation.

v The mix model, MC, offers the best combination because it gives the chance to make decisions inexpensive, and makes some specific corrections assuming an extra cost.

vi The team's constitution is strongly dependent on the cost of extra productivity; even when highly productive resources have a minor cost considering the USD/ton price for a rookie resource. If the fixed cost is high, the MS model prefers rookie resources. This is linked to the probability of recovery of the investment for this extra flexibility. This result contradicts the basic economical approach, where the value of the system is obtained without considering the potential realizations of uncertainty.

vii The methodology we used to determine the value of some decisions that are intuitive to managers, like teams' constitution and information buy, allows for the more scientific analysis of the alternatives.

viii RTA offers acceptable performances compared to MS and is always better compared to EV approach. The gap between RTA and MS increases when the secondary tree has minor periods of fan.

ix RTE has an acceptable performance for our set of temporal structures, independent of the rain probabilities, in profitable terms.

x RTW pays a high cost when the rain probability is low, because of its belief in allowing for large performance gaps.

xi The resources' ability diminishes the RTA-MS, being a critical issue.

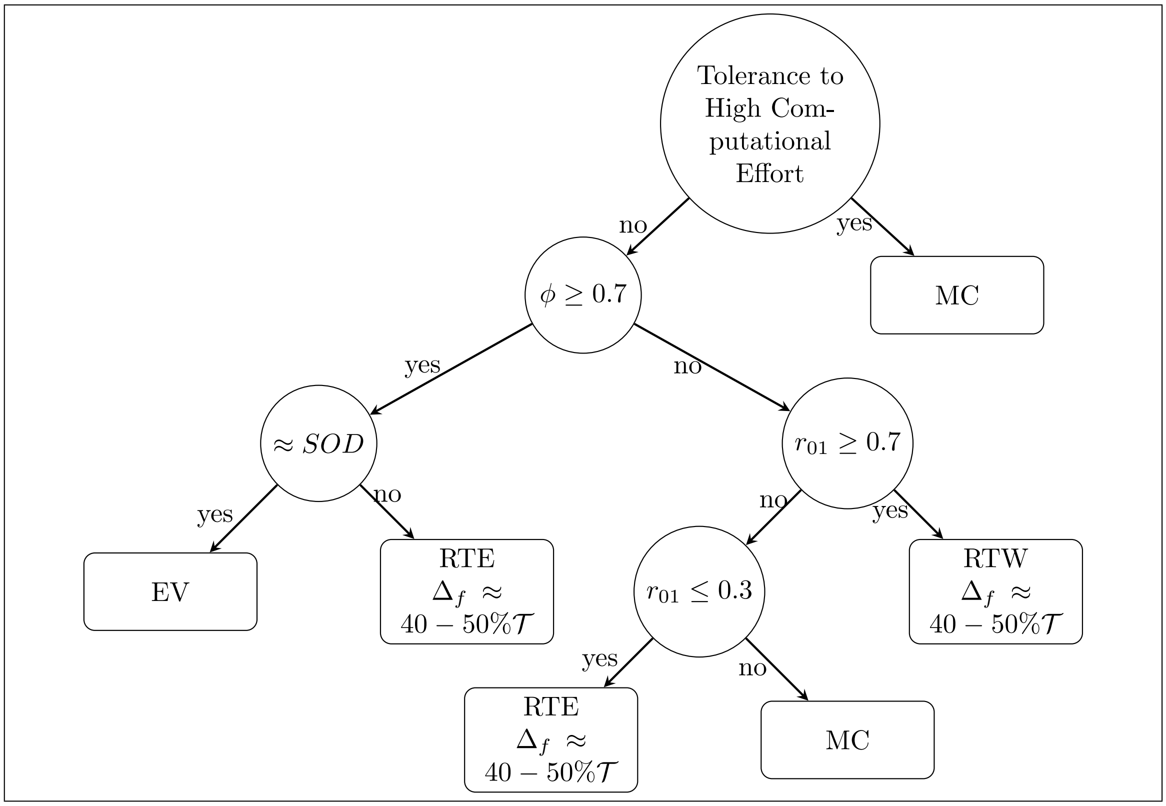



Figure 6.1. Quick tool to chose the model to apply according system features)

xii The computational time gain in using RTA instead of MS, is especially high when granularity in the secondary tree is low because it avoids the computational effort of big trees.

xiii We observed that the decisions trees are similar when the profit follows the same direction, but the general analysis indicates that there is a potential interaction with the time structure that requires to go deeper.

To conclude, the comparison of the models and conditions led to some criteria to choose the rights for different contexts of problems. Part of the facts that lead to the right choice of the model to use are presented in a flow diagram in Figure 6.1.

The first question is the tolerance of the decision-maker to a very high computational effort. If the decision-maker tolerates the effort, the variant with decisions in two epochs is preferred for the MC model. If the tolerance is low, we need to investigate other alternatives where the loss of performance is not more than 10% for the MC option. The first question in this branch is the ability of resource; if it is high, then the EV models could be good approximations. To decide which approach to use, the farmer should know about the dispersion of the optimum days of harvesting. If the optimum days are similar, the harvest will be concentrated, so the EV approach is sufficient. If the optimum days are dispersed, more sensibility is needed to anticipate or postpone harvesting; and RTE becomes a good option, specifically with a time structure that ensures granularity in the secondary tree for approximately 40–50% of the original tree. This granularity helps in capturing the versatility of the MS model while reducing the computational effort.

If the resource ability is medium or low, the farmer should analyze the weather forecast. For high rain probabilities, the RTW model with $\Delta_f \approx 40\text{–}50\%\mathcal{T}$, is acceptable. If the probability of rain is very low, RTE with $\Delta_f \approx 40\text{–}50\%\mathcal{T}$ is acceptable too. Lastly, the rain condition near the maximum uncertainty value, MC model, should be preferred, or an RTE with $\Delta_f \approx 40\text{–}50\%\mathcal{T}$ as a poorer alternative.

### 6.3. Future Work

This research requires going beyond the current scope, to deal with the interesting questions that arose throughout the development of the project, as well as with the relaxation of the assumptions that have simplified the real system. This section suggests some of these promising aspects:

i The flexibility price has been constant during the optimization time span. However, if more periods are considered, the market game resorts to variable prices. In that context, the decisions keep changing. If the models are incorporated with the new stochastic parameters, the size increases; thus, one feasible approach is to understand which interval allows the set decision to remain unaltered. Similar variability could be found in the real productivity of the resources, and even in other operational parameters.

ii There are biological aspects that could be improved; for example, the use of heat degree days, a more direct measure of the evolution of the fruit, which has even been modeled for financial derivatives that are not applied in optimization.

iii RTA offers positive signals in the first exploration. Risk aspects should be considered as a performance parameter and more experiments should be run to extend the conclusions.

iv The measured Nested Distance has proved the usefulness of having relative results; but it is difficult to understand what this means for the magnitude itself. A sensitivity study that is compared with performance indexes is useful to get knowledge in order to predict the structure of the decisions and information, validating the value of the system.

v The benefits of each model and the conditions in which they are to be applied need to be translated to the farmer language in their native language, thus, a *decision support system* (DSS) would be an interesting application to develop.

vi This work could be applied to other fresh fruits or crops that require a multi-period analysis to completely understand the impacts of practices and decisions.

vii This model, especially the RTA approach, offers an MS-simplified version but with a high potential to be applied in supply chain models that require the stochasticity without paying the computational cost.

# REFERENCES


Ahumada, O., and Villalobos, J. R. (2009). Application of planning models in the agri-food supply chain: A review. *European journal of Operational research*, *196*(1), 1–20.

Ahumada, O., and Villalobos, J. R. (2011). Operational model for planning the harvest and distribution of perishable agricultural products. *International Journal of Production Economics*, *133*(2), 677–687.

Ahumada, O., Villalobos], J. R., and Mason], A. N. (2012). Tactical planning of the production and distribution of fresh agricultural products under uncertainty. *Agricultural Systems*, *112*, 17 - 26. Retrieved from `http://www.sciencedirect.com/science/article/pii/S0308521X1200087X` doi: https://doi.org/10.1016/j.agsy.2012.06.002

Alexopoulos, K., Papakostas, N., Mourtzis, D., Gogos, P., and Chryssolouris, G. (2007). Quantifying the flexibility of a manufacturing system by applying the transfer function. *International Journal of Computer Integrated Manufacturing*, *20*(6), 538–547.

Allen, S. J., and Schuster, E. W. (2004). Controlling the risk for an agricultural harvest. *Manufacturing & Service Operations Management*, *6*(3), 225–236.

Alvarez de la Paz F., G. A., Reyes J. (n.d.). *Manual basico de viticultura en tacoronte-acentejo.* Retrieved 20.08.2021, from `http://www.tacovin.com/pdf/viti.pdf`

Arafa, A., and ElMaraghy, W. (2012). Quantifying the effect of enterprise strategic flexibility. In *Enabling manufacturing competitiveness and economic sustainability* (pp. 99–104). Springer.

Arnaout, J.-P. M., and Maatouk, M. (2010). Optimization of quality and operational costs through improved scheduling of harvest operations. *International Transactions in operational research*, *17*(5), 595–605.

Ashenfelter, O., Ashmore, D., and Lalonde, R. (1995). Bordeaux wine vintage quality and the weather. *Chance*, *8*(4), 7–14.

Banco Central de Chile. (n.d.). *Base de Datos Estadisticos* (Tech. Rep.). Accessed September 1, 2021 [Online]. Retrieved from `https://si3.bcentral.cl/Siete/ES/Siete/Cuadro/CAP_CCNN/MN_CCNN76/CCNN2013_IMACEC_01`

Barad, M. (2013). Flexibility development–a personal retrospective. *International Journal of Production Research*, *51*(23-24), 6803–6816.

Baykasoğlu, A. (2009). Quantifying machine flexibility. *International Journal of Production Research*, *47*(15), 4109–4123.

Beach, R., Muhlemann, A. P., Price, D. H., Paterson, A., and Sharp, J. A. (2000). A review of manufacturing flexibility. *European journal of operational research*, *122*(1), 41–57.

Beamon, B. M. (1999). Measuring supply chain performance. *International journal of operations & production management*.

Behzadi, G., O'Sullivan, M. J., Olsen, T. L., and Zhang, A. (2018). Agribusiness supply chain risk management: A review of quantitative decision models. *Omega*, *79*, 21–42.

Bertsimas, D., Gupta, V., and Kallus, N. (2018). Robust sample average approximation. *Mathematical Programming*, *171*(1), 217–282.

Besterfield, D. (2003). H., besterfield-michna, c., besterfield, g., h., besterfield-sacre, m. *Total quality management*, *3*.

Birge, J. R. (1982). The value of the stochastic solution in stochastic linear programs with fixed recourse. *Mathematical programming*, *24*(1), 314–325.

Birge, J. R., and Louveaux, F. (2011). *Introduction to stochastic programming*. Springer Science & Business Media.

Bischi, A., Taccari, L., Martelli, E., Amaldi, E., Manzolini, G., Silva, P., … Macchi, E. (2019). A rolling-horizon optimization algorithm for the long term operational scheduling of cogeneration systems. *Energy*, *184*, 73–90.

Bohle, C., Maturana, S., and Vera, J. (2010). A robust optimization approach to wine grape harvesting scheduling. *European Journal of Operational Research*, *200*(1), 245–252.

Borodin, V., Bourtembourg, J., Hnaien, F., and Labadie, N. (2014). A quality risk management problem: case of annual crop harvest scheduling. *International Journal of Production Research*, *52*(9), 2682–2695.

Borodin, V., Bourtembourg, J., Hnaien, F., and Labadie, N. (2016). Handling uncertainty in agricultural supply chain management: A state of the art. *European Journal of Operational Research*, *254*(2), 348–359.

Boulton, R. B., Singleton, V. L., Bisson, L. F., and Kunkee, R. E. (2013). *Principles and practices of winemaking*. Springer Science & Business Media.

Buzacott, J. A., and Mandelbaum, M. (2008). Flexibility in manufacturing and services: achievements, insights and challenges. *Flexible Services and Manufacturing Journal*, *20*(1-2), 13.

Cardin, M.-A., and Hu, J. (2016). Analyzing the tradeoffs between economies of scale, time-value of money, and flexibility in design under uncertainty: Study of centralized versus decentralized waste-to-energy systems. *Journal of Mechanical Design*, *138*(1), 11401.

*Castastro Viticola Nacional 2019 vinos de chile.* (n.d.). `https://www.odepa.gob.cl/rubro/vinos/catastro-viticola-nacional`. (Accessed: 2021-07-01)

Champion, B. R., and Gabriel, S. A. (2017). A multistage stochastic energy model with endogenous probabilities and a rolling horizon. *Energy and Buildings*, *135*, 338–349.

Chand, S., Hsu, V. N., and Sethi, S. (2002). Forecast, solution, and rolling horizons in operations management problems: A classified bibliography. *Manufacturing & Service Operations Management*, *4*(1), 25–43.

Chen, X., Kouvelis, P., and Biazaran, M. (2018). Value of operational flexibility in co-production systems with yield and demand uncertainty. *International Journal of Production Research*, *56*(1-2), 491–507.

Cholette, S. (2009). Mitigating demand uncertainty across a winery's sales channels through postponement. *International Journal of Production Research*, *47*(13), 3587–3609.

Chryssolouris, G., Efthymiou, K., Papakostas, N., Mourtzis, D., and Pagoropoulos, A. (2013). Flexibility and complexity: is it a trade-off? *International Journal of Production Research*, *51*(23-24), 6788–6802.

Coombe, B. (1992). Research on development and ripening of the grape berry. *American Journal of Enology and Viticulture*, *43*(1), 101–110.

Cuturi, M. (2013). Sinkhorn distances: Lightspeed computation of optimal transport. *Advances in neural information processing systems*, *26*, 2292–2300.

Dai, Z., and Li, Y. (2013). A multistage irrigation water allocation model for agricultural land-use planning under uncertainty. *Agricultural water management*, *129*, 69–79.

Das, A., and Caprihan, R. (2008). A rule-based fuzzy-logic approach for the measurement of manufacturing flexibility. *The International Journal of Advanced Manufacturing Technology*, *38*(11), 1098–1113.

De Toni, A., and Tonchia, S. (1998). Manufacturing flexibility: a literature review. *International journal of production research*, *36*(6), 1587–1617.

Devine, M. T., Gabriel, S. A., and Moryadee, S. (2016). A rolling horizon approach for stochastic mixed complementarity problems with endogenous learning: Application to natural gas markets. *Computers & Operations Research*, *68*, 1–15.

Devine, M. T., Gleeson, J. P., Kinsella, J., and Ramsey, D. M. (2014). A rolling optimisation model of the uk natural gas market. *Networks and Spatial Economics*, *14*(2), 209–244.

Dupačová, J., Gröwe-Kuska, N., and Römisch, W. (2003). Scenario reduction in stochastic programming. *Mathematical programming*, *95*(3), 493–511.

Escudero, L. F., Garín, A., Merino, M., and Pérez, G. (2007). The value of the stochastic solution in multistage problems. *Top*, *15*(1), 48–64.

Esmaeilikia, M., Fahimnia, B., Sarkis, J., Govindan, K., Kumar, A., and Mo, J. (2016). Tactical supply chain planning models with inherent flexibility: definition and review. *Annals of Operations Research*, *244*(2), 407–427.

Esteso, A., Alemany, M. M., and Ortiz, A. (2018). Conceptual framework for designing agri-food supply chains under uncertainty by mathematical programming models. *International Journal of Production Research*, *56*(13), 4418–4446.

Esturilho, C. G., and Estorilio, C. (2010). The deployment of manufacturing flexibility as a function of company strategy. *Journal of Manufacturing Technology Management*.

Falcão, L. D., Chaves, E. S., Burin, V. M., Falcão, A. P., Gris, E. F., Bonin, V., and Bordignon-Luiz, M. T. (2008). Maturity of cabernet sauvignon berries from grapevines grown with two different training systems in a new grape growing region in brazil. *Ciencia e investigación agraria*, *35*(3), 321–332.

Ferrer, J.-C., Mac Cawley, A., Maturana, S., Toloza, S., and Vera, J. (2008). An optimization approach for scheduling wine grape harvest operations. *International Journal of Production Economics*, *112*(2), 985–999.

Flamary, R., Courty, N., Gramfort, A., Alaya, M. Z., Boisbunon, A., Chambon, S., ... Vayer, T. (2021). Pot: Python optimal transport. *Journal of Machine Learning Research*, *22*(78), 1-8. Retrieved from `http://jmlr.org/papers/v22/20-451.html`

Glories, Y., Ribéreau-Gayon, P., Maujean, A., and Dubourdieu, D. (2000). *Handbook of enology: The chemistry of wine: Stabilization and treatments*. John Wiley & Sons.

Goyal, M., and Netessine, S. (2007). Strategic technology choice and capacity investment under demand uncertainty. *Management Science*, *53*(2), 192–207.

Guan, Z., and Philpott, A. B. (2011). A multistage stochastic programming model for the new zealand dairy industry. *International Journal of Production Economics*, *134*(2), 289–299.

Gupta, D. (1993). On measurement and valuation of manufacturing flexibility. *The International Journal of Production Research*, *31*(12), 2947–2958.

Gupta, D., and Buzacott, J. A. (1996). A "goodness test" for operational measures of manufacturing flexibility. *International Journal of Flexible Manufacturing Systems*, *8*(3), 233–245.

Gupta, Y. P., and Somers, T. M. (1996). Business strategy, manufacturing flexibility, and organizational performance relationships: a path analysis approach. *Production and Operations Management*, *5*(3), 204–233.

Haeger, J. W., and Storchmann, K. (2006). Prices of american pinot noir wines: climate, craftsmanship, critics. *Agricultural economics*, *35*(1), 67–78.

Hart, W. E., Laird, C. D., Watson, J.-P., Woodruff, D. L., Hackebeil, G. A., Nicholson, B. L., and Siirola, J. D. (2017). *Pyomo-optimization modeling in python* (Vol. 67). Springer.

Heitsch, H., and Römisch, W. (2009). Scenario tree modeling for multistage stochastic programs. *Mathematical Programming*, *118*(2), 371–406.

Heitsch, H., Römisch, W., and Strugarek, C. (2006). Stability of multistage stochastic programs. *SIAM Journal on Optimization*, *17*(2), 511–525.

Higle, J. L., and Wallace, S. W. (2003). Sensitivity analysis and uncertainty in linear programming. *Interfaces*, *33*(4), 53–60.

Horejšová, M., Vitali, S., Kopa, M., and Moriggia, V. (2020). Evaluation of scenario reduction algorithms with nested distance. *Computational Management Science*, *17*(2), 241–275.

Høyland, K., and Wallace, S. W. (2001). Generating scenario trees for multistage decision problems. *Management science*, *47*(2), 295–307.

Hu, Z., and Hu, G. (2018). A multi-stage stochastic programming for lot-sizing and scheduling under demand uncertainty. *Computers & Industrial Engineering*, *119*, 157 - 166.

Huang, G. H., and Loucks, D. P. (2000). An inexact two-stage stochastic programming model for water resources management under uncertainty. *Civil Engineering and Environmental Systems*, *17*(2), 95-118.

Huang, K., and Ahmed, S. (2009). The value of multistage stochastic programming in capacity planning under uncertainty. *Operations Research*, *57*(4), 893–904.

Huh, W. T., and Lall, U. (2013). Optimal crop choice, irrigation allocation, and the impact of contract farming. *Production and Operations Management*, *22*(5), 1126–1143.

International Organisation of Vine and Wine. (n.d.). *2019 Statistical Report on World Vitiviniculture* (Tech. Rep.). Accessed July 20, 2021 [Online]. Retrieved from `https://www.oiv.int/public/medias/6782/oiv-2019-statistical-report-on-world-vitiviniculture.pdf`

Jackson, R. S. (2008). *Wine science: principles and applications*. Academic press.

Jain, A., Jain, P., Chan, F. T., and Singh, S. (2013). A review on manufacturing flexibility. *International Journal of Production Research*, *51*(19), 5946–5970.

Jones, G. V., White, M. A., Cooper, O. R., and Storchmann, K. (2005). Climate change and global wine quality. *Climatic change*, *73*(3), 319–343.

Jonkman, J., Barbosa-Póvoa, A. P., and Bloemhof, J. M. (2019). Integrating harvesting decisions in the design of agro-food supply chains. *European Journal of Operational Research*, *276*(1), 247–258.

Jordan, W. C., and Graves, S. C. (1995). Principles on the benefits of manufacturing process flexibility. *Management science*, *41*(4), 577–594.

Kahyaoglu, Y., Kayaligil, S., and Buzacott, J. (2002). Flexibility analysis: a methodology and a case study. *International journal of production research*, *40*(16), 4111–4130.

Kazemi Zanjani, M., Nourelfath, M., and Ait-Kadi, D. (2010). A multi-stage stochastic programming approach for production planning with uncertainty in the quality of raw materials and demand. *International Journal of Production Research*, *48*(16), 4701–4723.

Kazemi Zanjani, M., Nourelfath, M., and Aït-Kadi, D. (2011). Production planning with uncertainty in the quality of raw materials: a case in sawmills. *Journal of the Operational Research Society*, *62*(7), 1334–1343.

Kennedy, J. O. (2012). *Dynamic programming: applications to agriculture and natural resources*. Springer Science & Business Media.

Kim, S., Pasupathy, R., and Henderson, S. G. (2015). A guide to sample average approximation. *Handbook of simulation optimization*, 207–243.

Kleywegt, A. J., Shapiro, A., and Homem-de Mello, T. (2002). The sample average approximation method for stochastic discrete optimization. *SIAM Journal on Optimization*, *12*(2), 479–502.

Koste, L. L., and Malhotra, M. K. (1999). A theoretical framework for analyzing the dimensions of manufacturing flexibility. *Journal of operations management*, *18*(1), 75–93.

Koste, L. L., Malhotra, M. K., and Sharma, S. (2004). Measuring dimensions of manufacturing flexibility. *Journal of Operations Management*, *22*(2), 171–196.

Kovacevic, R. M., and Pichler, A. (2015). Tree approximation for discrete time stochastic processes: a process distance approach. *Annals of operations research*, *235*(1), 395–421.

Kusumastuti, R. D., van Donk, D. P., and Teunter, R. (2016). Crop-related harvesting and processing planning: a review. *International Journal of Production Economics*, *174*, 76–92.

Labrianidis, L. (1995). Flexibility in production through subcontracting: the case of the poultry meat industry in greece. *Environment and Planning A*, *27*(2), 193–209.

Le Moigne, M., Symoneaux, R., and Jourjon, F. (2008). How to follow grape maturity for wine professionals with a seasonal judge training? *Food Quality and Preference*, *19*(8), 672–681.

Li, Q., and Hu, G. (2020). Multistage stochastic programming modeling for farmland irrigation management under uncertainty. *Plos one*, *15*(6), e0233723.

Li, W., Li, Y., Li, C., and Huang, G. (2010). An inexact two-stage water management model for planning agricultural irrigation under uncertainty. *Agricultural Water Management*, *97*(11), 1905 - 1914.

Lima, J.L. (2015). *Estudio de caracterización de la cadena de producción y comercialización de la agroindustria vitivinícola: estructura, agentes y practicas* (Tech. Rep.). Santiago, Chile: ODEPA. Retrieved from https://www.odepa.gob.cl/wp-content/uploads/2017/12/AgroindustriaVitivinicola.pdf

Lobos, A., and Vera, J. R. (2016). Intertemporal stochastic sawmill planning: Modeling and managerial insights. *Computers & Industrial Engineering*, *95*, 53–63.

Lund, S. T., and Bohlmann, J. (2006). The molecular basis for wine grape quality-a volatile subject. *Science*, *311*(5762), 804–805.

Madansky, A. (1959). *Some results and problems in stochastic linear programming* (Tech. Rep.). RAND CORP SANTA MONICA CA.

Maggioni, F., Allevi, E., and Bertocchi, M. (2014). Bounds in multistage linear stochastic programming. *Journal of Optimization Theory and Applications*, *163*(1), 200–229.

Maggioni, F., and Wallace, S. W. (2012). Analyzing the quality of the expected value solution in stochastic programming. *Annals of Operations Research*, *200*(1), 37–54.

Mandelbaum, M., and Buzacott, J. (1990). Flexibility and decision making. *European Journal of Operational Research*, *44*(1), 17–27.

Meléndez, E., Ortiz, M., Sarabia, L., Íñiguez, M., and Puras, P. (2013). Modelling phenolic and technological maturities of grapes by means of the multivariate relation between organoleptic and physicochemical properties. *Analytica Chimica Acta*, *761*, 53–61.

*Memoria Anual 2019 vinos de chile.* (n.d.). `https://www.winesofchile.org/wp-content/uploads/2020/07/MEMORIA-WOC-FINAL-WEB.pdf`. (Accessed: 2021-07-19)

Mezgár, I., Kovács, G. L., and Paganelli, P. (2000). Co-operative production planning for small-and medium-sized enterprises. *International Journal of Production Economics*, *64*(1-3), 37–48.

Mishra, R., Pundir, A. K., and Ganapathy, L. (2014). Manufacturing flexibility research: A review of literature and agenda for future research. *Global Journal of Flexible Systems Management*, *15*(2), 101–112.

Moghaddam, K. S., and DePuy, G. W. (2011). Farm management optimization using chance constrained programming method. *Computers and electronics in agriculture*, *77*(2), 229–237.

Morata, A. (2018). *Red wine technology*. Academic Press.

Morton, T. E. (1981). Forward algorithms for forward thinking managers. *Applications of Management Science*, *1*, 1–55.

Mulvey, J. M., and Ruszczyński, A. (1995). A new scenario decomposition method for large-scale stochastic optimization. *Operations research*, *43*(3), 477–490.

Pagnoncelli, B. K., Ahmed, S., and Shapiro, A. (2009). Sample average approximation method for chance constrained programming: theory and applications. *Journal of optimization theory and applications*, *142*(2), 399–416.

Pantuso, G., and Boomsma, T. K. (2019). On the number of stages in multistage stochastic programs. *Annals of Operations Research*, 1–23.

Patel, P. C. (2011). Role of manufacturing flexibility in managing duality of formalization and environmental uncertainty in emerging firms. *Journal of Operations Management*, *29*(1-2), 143–162.

Patel, P. C., Terjesen, S., and Li, D. (2012). Enhancing effects of manufacturing flexibility through operational absorptive capacity and operational ambidexterity. *Journal of*

*Operations Management*, *30*(3), 201–220.

Pflug, G. C. (2010). Version-independence and nested distributions in multistage stochastic optimization. *SIAM Journal on Optimization*, *20*(3), 1406–1420.

Pflug, G. C., and Pichler, A. (2016). *Multistage stochastic optimization*. Springer.

Ramos, M., Jones, G., and Martínez-Casasnovas, J. (2008). Structure and trends in climate parameters affecting winegrape production in northeast spain. *Climate Research*, *38*(1), 1–15.

Rardin, R. L. (2016). *Optimization in operations research. 2. utg*. Boston: Pearson.

Reynolds, A. G. (2010). *Managing wine quality: viticulture and wine quality*. Elsevier.

Richardson, C. W., and Wright, D. A. (1984). Wgen: A model for generating daily weather variables. *ARS (USA)*.

Rockafellar, R. T., and Wets, R. J.-B. (1991). Scenarios and policy aggregation in optimization under uncertainty. *Mathematics of operations research*, *16*(1), 119–147.

Rogalski, S. (2011). *Flexibility measurement in production systems*. Springer.

Rong, A., Akkerman, R., and Grunow, M. (2011). An optimization approach for managing fresh food quality throughout the supply chain. *International Journal of Production Economics*, *131*(1), 421–429.

Roohnavazfar, M., Manerba, D., De Martin, J. C., and Tadei, R. (2019). Optimal paths in multi-stage stochastic decision networks. *Operations Research Perspectives*, *6*, 100124.

Ruszczyński, A., and Shapiro, A. (2003). Stochastic programming models. *Handbooks in operations research and management science*, *10*, 1–64.

Sawhney, R. (2006). Interplay between uncertainty and flexibility across the value-chain: towards a transformation model of manufacturing flexibility. *Journal of operations management*, *24*(5), 476–493.

Sethi, A. K., and Sethi, S. P. (1990). Flexibility in manufacturing: a survey. *International journal of flexible manufacturing systems*, *2*(4), 289–328.

Sethi, S., and Sorger, G. (1991). A theory of rolling horizon decision making. *Annals of Operations Research*, *29*(1), 387–415.

Seyoum-Tegegn, E., and Chan, C. (2013). What is making vineyard investment in north-west victoria, australia, slow to adjust? *Journal of Wine Economics*, *8*(1), 83.

Shapiro, A. (2010). Computational complexity of stochastic programming: Monte carlo sampling approach. In *Proceedings of the international congress of mathematicians 2010 (icm 2010) (in 4 volumes) vol. i: Plenary lectures and ceremonies vols. ii–iv: Invited lectures* (pp. 2979–2995).

Shapiro, A., and Nemirovski, A. (2005). On complexity of stochastic programming problems. In *Continuous optimization* (pp. 111–146). Springer.

Shi, D., and Daniels, R. L. (2003). A survey of manufacturing flexibility: Implications for e-business flexibility. *IBM Systems Journal*, *42*(3), 414–427.

Silvente, J., Kopanos, G. M., Dua, V., and Papageorgiou, L. G. (2018). A rolling horizon approach for optimal management of microgrids under stochastic uncertainty. *Chemical Engineering Research and Design*, *131*, 293–317.

Slack, N. (1983). Flexibility as a manufacturing objective. *International Journal of Operations & Production Management*.

Soto-Silva, W. E., Nadal-Roig, E., González-Araya, M. C., and Pla-Aragones, L. M. (2016). Operational research models applied to the fresh fruit supply chain. *European Journal of Operational Research*, *251*(2), 345–355.

Swafford, P. M., Ghosh, S., and Murthy, N. (2006). The antecedents of supply chain agility of a firm: scale development and model testing. *Journal of Operations management*, *24*(2), 170–188.

Tadei, R., Perboli, G., and Manerba, D. (2019). The multi-stage dynamic stochastic decision process with unknown distribution of the random utilities. *Optimization Letters*, 1–12.

Terkaj, W., Tolio, T., and Valente, A. (2009). A review on manufacturing flexibility. *Design of flexible production systems*, 41–61.

Thapalia, B. K., Wallace, S. W., and Kaut, M. (2009). Using inventory to handle risks in the supply of oil to nepal. *International Journal of Business Performance and Supply Chain Modelling*, *1*(1), 41–60.

Timonina, A. V. (2015). Multi-stage stochastic optimization: the distance between stochastic scenario processes. *Computational Management Science*, *12*(1), 171–195.

Upton, D. M. (1994). The management of manufacturing flexibility. *California management review*, *36*(2), 72–89.

Urdiales, D., Meza, F., Gironás, J., and Gilabert, H. (2018). Improving stochastic modelling of daily rainfall using the enso index: Model development and application in chile. *Water*, *10*(2), 145.

Van Der Vorst, J. G., Tromp, S.-O., and Zee, D.-J. v. d. (2009). Simulation modelling for food supply chain redesign; integrated decision making on product quality, sustainability and logistics. *International Journal of Production Research*, *47*(23), 6611–6631.

van Hop, N., and Ruengsak, K. (2005). Fuzzy estimation for manufacturing flexibility. *International Journal of Production Research*, *43*(17), 3605–3617.

van Leeuwen, C., and Darriet, P. (2016). The impact of climate change on viticulture and wine quality. *Journal of Wine Economics*, *11*(1), 150–167.

Varas, M., Maturana, S., Cholette, S., Mac Cawley, A., and Basso, F. (2018). Assessing the benefits of labelling postponement in an export-focused winery. *International Journal of Production Research*, *56*(12), 4132–4151.

Veliz, F. B., Watson, J.-P., Weintraub, A., Wets, R. J.-B., and Woodruff, D. L. (2015). Stochastic optimization models in forest planning: a progressive hedging solution approach. *Annals of Operations Research*, *232*(1), 259–274.

Villani, C. (2003). *Topics in optimal transportation* (No. 58). American Mathematical Soc.

Vitali, S. (2018). Multistage multivariate nested distance: an empirical analysis. *Kybernetika*, *54*(6), 1184–1200.

Wang, T. (2005). *Real options "in" projects and systems design : identification of options and solutions for path dependency* (PhD dissertation). Massachusetts Institute of Technology.

Wang, W., and Ahmed, S. (2008). Sample average approximation of expected value constrained stochastic programs. *Operations Research Letters*, *36*(5), 515–519.

Wiedenmann, S., and Geldermann, J. (2015). Supply planning for processors of agricultural raw materials. *European Journal of Operational Research*, *242*(2), 606–619.

Zhang, C., Li, M., and Guo, P. (2017). An interval multistage joint-probabilistic chance-constrained programming model with left-hand-side randomness for crop area planning under uncertainty. *Journal of Cleaner Production*, *167*, 1276–1289.

Zhang, F., Guo, P., Engel, B. A., Guo, S., Zhang, C., and Tang, Y. (2019). Planning seasonal irrigation water allocation based on an interval multiobjective multi-stage stochastic programming approach. *Agricultural Water Management*, *223*, 105692.

Zhang, H., Ha, M., Zhao, H., and Song, J. (2017). Inexact multistage stochastic chance constrained programming model for water resources management under uncertainties. *Scientific Programming*, *2017*.

Zhou, Y., Huang, G. H., and Yang, B. (2013). Water resources management under multi-parameter interactions: A factorial multi-stage stochastic programming approach. *Omega*, *41*(3), 559 - 573.

# APPENDIX

# A. WINE: MARKET, NATURE AND INDUSTRIALIZATION

## Wine Sector in Chile

In figure A.1 world wine annual production by country is shown. The traditional winemakers (France, Italy, and Spain) contribute to the total output, setting aside Chile, Australia, and Argentina, the new players in the international context.

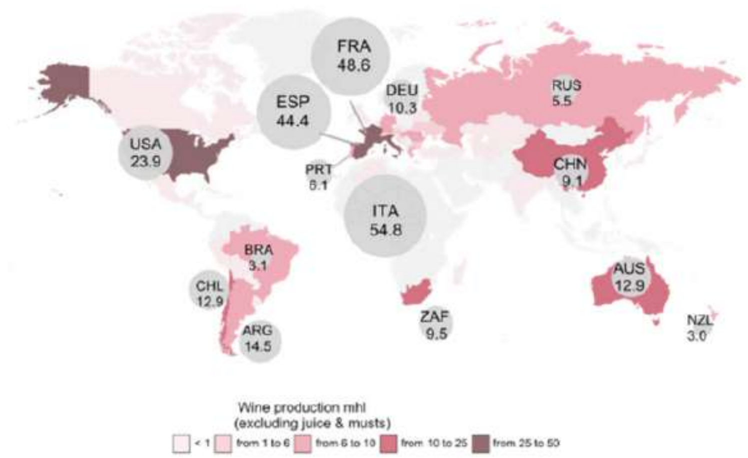



Figure A.1. Wine Production in 2018 in different countries. Source: International Organisation of Vine and Wine (n.d.)

However, although Chile is peripheral in terms of production, it plays a critical role in exportations. In table A.1, Chile reaches the four position at the global level with 9% of the total exported volume.

In figure A.2, the evolution of exported volumes is shown for the Chilean case. There are two different groups of exportation, depending on the appellation of origin certificate (AO). Technically, the wines with an AO are from specific regions established by ministerial decree that reaches certain standards. For example, a wine with the *Valle de Maule cabernet sauvignon, 2015* AO, requires that 75 % of the grapes come from Maule Valley, 75% of the grape should be Cabernet Sauvignon grape, and it has to be produced in 2015 in the same percentage. Different abbreviations are used to refer to this feature; see *https://www.wine-searcher.com/wine-terms* for more details. The wine without this feature

Table A.1. Main exporters in millions of hl. Source:International Organisation of Vine and Wine (n.d.)

| **Country** | **2014** | **2015** | **2016** | **2017** | **2018** |
|---|---|---|---|---|---|
| Spain | 22.0% | 23.0% | 22.0% | 21.0% | 20.0% |
| Italy | 20.0% | 19.0% | 20.0% | 20.0% | 18.0% |
| France | 14.0% | 13.0% | 14.0% | 14.0% | 13.0% |
| Chile | 8.0% | 8.0% | 9.0% | 9.0% | 9.0% |
| Australia | 7.0% | 7.0% | 7.0% | 7.0% | 8.0% |
| South Africa | 4.0% | 4.0% | 4.0% | 4.0% | 5.0% |
| Germany | 4.0% | 4.0% | 3.0% | 4.0% | 3.0% |
| USA | 4.0% | 4.0% | 4.0% | 3.0% | 3.0% |
| Portugal | 3.0% | 3.0% | 3.0% | 3.0% | 3.0% |
| Argentina | 3.0% | 3.0% | 3.0% | 2.0% | 3.0% |
| New Zealand | 2.0% | 2.0% | 2.0% | 2.0% | 2.0% |
| Moldova | 1.0% | 1.0% | 1.0% | 1.0% | 1.0% |

is bulk wine. The total volume of chilean exports has increased in the last twenty years, with a drop in the previous three years. AO and bulk exportations follow the same patterns. It's important to highlight that the AO volume is around 15% bigger than the bulk ones because AOPs have more added value when exporting, which is created in Chile. In figure A.3 the average exportation price is shown for 2000-2020 period. Prices present a slight positive trend, but there are generally stables, which are similar for both types of exportations. The AOP price is up to 6 times the bulk price, but the costs are also higher. According to **?**, China is the leading destination with 15% of the annual volume of the DO wine, followed by Japan, United Kingdom, and Brazil, with around 11% each of them. These four countries represent almost the 50% in monetary terms. The average price expressed in usd/l FOB is 3.4, 2.7, 2.9, and 2.5, for China, Japan, UK, and Brazil, respectively. The grannel business is smaller than the DO, and USA and Argentina are the main destinies, meaning 30% of the total annual volume. In figure **??**, appears the exportation bottled volume grouped by the price of each box. The figure goes through different years.

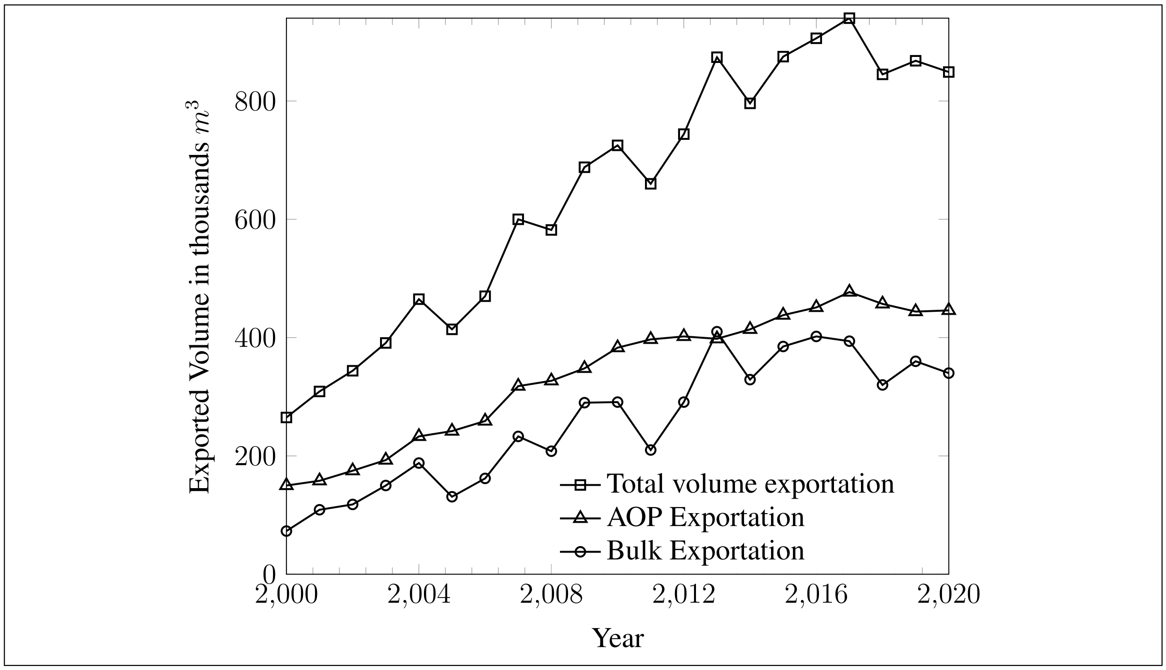



Figure A.2. Final annual volume exported in periods 2000-2020

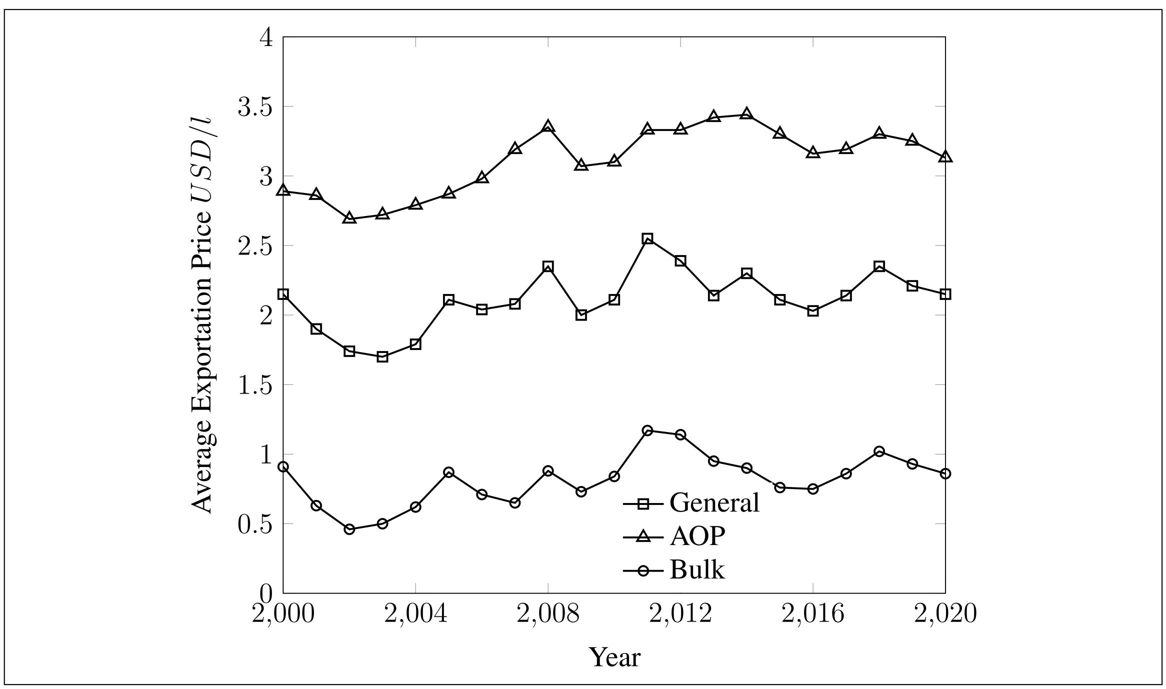



Figure A.3. Average exportation price in periods 2000-2020

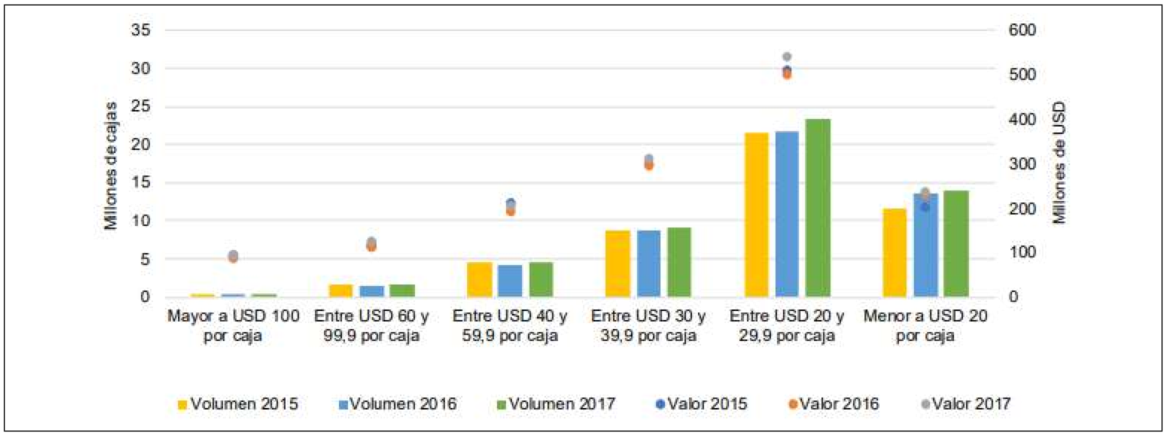



Figure A.4. Bottled Wine exportation gathered by price range. Source: (International Organisation of Vine and Wine, n.d.)

Besides the international market, *Memoria Anual 2019 Vinos de Chile* (n.d.) indicates that inner wine consumption per capita yearly is 14.1 liters (2019). The local market size reaches almost USD 1,000 million, with an annual volume of 248.4 million liters. As we mentioned before, the sector's contribution to the GDP is important. It plays a fundamental role in employment and local economic development because of the vast region of the crops. According to *Castastro Viticola Nacional 2019 Vinos de Chile* (n.d.), in 2019, 136,000 hectares of wine grapes are reached, with 26 % of white strains; the distribution is shown in table A.2.

The surface growth has been sustained over time and almost tripled the 1995 production. The investment requires at least 4-5 years to produce with full capacityLima, J.L. (2015). An extra cost that reports better yields is the irrigation systems, which promotes the opportunity to explore new geographical zones for production. The initial prices of the plantations represent in the order of 20 % of the total cost of the vineyard's life, estimated in 20 years for accounting use, but with an actual life span around 30-40 years. Grape costs are significant in this sector because of the variability of the prices. In table A.3 15 years of history of prices for some varieties are shown, and the variability among years and harvest moment is evident, giving significant weight to the management capacity to make profitable the cultivars.

The bottled wine price differs considerably from the price paid to producers. The supply chain of the wine explains part of this gap. There are four clearly defined participants:

Table A.2. Vineyard land distribution in Chile in 2019 (*Castastro Viticola Nacional 2019 Vinos de Chile*, n.d.)

| Region | Area (ha) | **White strains** | **Red strains** |
|---|---|---|---|
| ARICA | 15 | | 15.00 |
| TARAPACA | 3.10 | 1.30 | 1.80 |
| ANTOFAGASTA | 4.97 | 1.06 | 3.91 |
| DE ATACAMA | 48.62 | 21.43 | 27.19 |
| DE COQUIMBO | 3,147.55 | 1,784.28 | 1,363.27 |
| DE VALPARAISO | 9,657.20 | 6,251.63 | 3,405.57 |
| DEL L.G.B. O'HIGGINS | 45,142.42 | 6,545.80 | 38,596.62 |
| DEL MAULE | 53,818.68 | 14,290.95 | 39,527.73 |
| ÑUBLE | 10,172.21 | 4244.13 | 5928.08 |
| DEL BIO BIO | 2,581.87 | 1300.45 | 1281.42 |
| DE LA ARAUCANIA | 84.55 | 38.69 | 45.86 |
| LOS RIOS | 18.50 | 13.70 | 4.80 |
| DE LOS LAGOS | 9.25 | 2.59 | 6.66 |
| METROPOLITANA DE SANTIAGO | 11,584.87 | 1,428.80 | 10,156.07 |
| **TOTAL** | **136,288.79** | **35,924.81** | **100,363.98** |

grape producers, collectors/intermediaries, wine producers, and wine marketers. The form of commercialization in the Chilean market includes three different ways of contracting:

(i) Long Term: they are expected when the arable land is of quality for super-premium wines. The winemaker company establishes a relationship with the producer, where the interference extends to agricultural practices. As a disadvantage, this type of contract requires permanent monitoring, and the prices paid per kilogram of grape are high compared to other options.

(ii) Annual: in this contract, a range of tonnes is received, and a base price is set, corrected by quality parameters.

(iii) Spot Market: there is no contract between the producer and the wine producer before the harvest event. The intermediaries play the role of negotiators, agreeing on prices based on the demands and conditions of the grape to be purchased.

Table A.3. Producer nominal price in Chilean pesos per kilogram of harvested grape. Nuble region, periods: 2002-2003 to 2016-2017- *Fuente: Seremi de Agricultura Región del Bío Bío.*

| | Harvest Beginning | | Harvest End | |
|---|---|---|---|---|
| Year | País | Moscatel de Alejandría | País | Moscatel de Alejandría |
| 2002-2003 | 45-50 | 50-70 | 55-60 | 60-70 |
| 2003-2004 | 65 | 75 | 85 | 80 |
| 2004-2005 | 135 | 140 | 110 | 120 |
| 2005-2006 | 50 | 60 | 50 | 60 |
| 2006-2007 | 25-35 | 40-60 | 25-35 | 25-30 |
| 2007-2008 | 70 | 70 | 70-80 | 70-75 |
| 2008-2009 | 50 | 50 | 50 | 50 |
| 2009-2010 | 100 | 100 | 120-130 | 120 |
| 2010-2011 | 150 | 150 | 180 | 180 |
| 2011-2012 | 130 | 130-150 | 100-130 | 100-130 |
| 2012-2013 | 80 | 80 | 90 | 100 |
| 2013-2014 | 80 | 120 | 100 | 140 |
| 2014-2015 | 70 | 85 | 60 | 70 |
| 2015-2016 | 75 | 90 | 85 | 100 |
| 2016-2017 | 130 | 100 | 190 | 165 |

The contractual strategies should aim to maintain the quality of the wine and a low cost for wine producers. Usually, a portfolio of contracts is kept, including spot market volume, to adjust some engagement volume variability or to take advantage of overproduction

Table A.4. Main activities in farm and months when they are done (based on Alvarez de la Paz F. (18.10.2005/n.d.)

| | Months | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| North | Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec |
| South | Jul | Aug | Sep | Oct | Nov | Dec | Jan | Feb | Mar | Apr | May | Jun |
| Land tillage | | | | ■ | | | | | | ■ | | |
| Weed control | ■ | | ■ | | ■ | | | | | | | |
| Prunning | ■ | ■ | ■ | | | | | | | | | |
| Phytosanitary application | | | | ■ | ■ | ■ | ■ | | | | | ■ |
| Sulfur added | | | | ■ | ■ | ■ | | | | | | |
| Defoliate | | | | | | ■ | | | | | | |
| Green Prunning | | | | | ■ | ■ | | | | | | |
| Irrigation | | | ■ | ■ | ■ | ■ | | ■ | ■ | ■ | | |
| Fertilization | | | ■ | ■ | ■ | ■ | | ■ | ■ | ■ | | |
| Grape Harvest | | | | | | | | ■ | ■ | ■ | | |

opportunities.

***Vitis Vinifera***

*Vitis vinifera* quality is critical to obtain a qualified wine. In this section, we briefly discuss the farm tasks to end in the harvesting process. Next, we describe the ripening curve and the weather impact qualitatively.

The *Vitis vinifera* planning includes several steps that are cyclic as the annual tasks, but others are fundamental, like the soil choice, and could be done only once. The potential quality of the grape is influenced by the soil selection, weather, and operations during the plant's shelf life. Soil and operations are decisions that could be controlled to some degree, diminishing the uncertainty. Still, weather is essentially uncertain, unstable through the years, and additionally, its weight is high for quality and yield production.

In table A.4 we present the gathered tasks in the farm (Alvarez de la Paz F., 18.10.2005/n.d.), and when they are done considering the months of the year, corrected by the hemisphere where the production is carried on.

The initial step is the selection of the soil. The choice considers some drivers: good drainage, protection from winds, and the type of biodiversity present. The level of soil drainage is critical since excess water has positive effects on the growth and vigor of the plant but not on the development of the grape; This flexibility comes from the type of soil but also from its management. After leveling the soil according to the type of irrigation chosen, pre-planting activities like subsoiling (consisting of breaking the ground to facilitate the growth of the roots and the retention of water) and fertilization are carried out. It usually consists of increasing the organic matter of the soil and a correction of essential elements.

The planting stage is itself critical. Momentum, varieties, and training system are defined. The latter refers to how the growth of the vineyard branches will be conducted and will influence the life of the vineyard because important tasks like pruning and harvesting depend on that. The training system is also responsible for the planting density; around 5,000 to 10,000 plants/hectare is adequate for quality wines.

Once the plant is set, it is necessary to guide the efforts to product development, and the pruning step is one of the critical stages. Pruning refers to the removal of plant organs, shoots, leaves, clusters, and others. It impacts in quality and yield of the grape and is frequently used as a milestone to determine the moment of maximum production. Biologically, pruning limits the energy destined for non-productive purposes. Therefore it allows for balancing the conditions of the season with the objective of the plantation. There is a winter and a summer pruning; the first is orientated to the maintenance of the plant and the second for productive purposes (carried out in the vegetation stage).

Fertilization also plays a vital role throughout the life of the plant. It seeks to restore the capacities of the soil, given the wear it has due to consumption and helps with micro components that correct some deficiencies. Irrigation is another critical practice because the water regime needs to be such that it avoids the stress that will result in production losses.

All the previous efforts end in the harvest stage, the step that we will address in this work in greater detail. It involves determining the degree of maturity, the removal of clusters, and the transport to the industry. The quality that we harvest is the maximum that the wine will achieve, so the moment and status of the grape should be chosen carefully. Harvesting can be manual, mechanized, or dually. The decision depends on several issues, variety, the winery reception program, weather, capacity of the vineyard, and availability of labor. The manual harvest is versatile in terms of vineyard driving; you have to observe the ergonomic limitations. In automatic harvesting, the driving system is limiting, and it is necessary to think about the crop's development with this orientation. The advantage is that it can operate 24 hours a day, which gives it more sensitivity to harvest at times where low temperatures preserve quality.

The flexibility that each option adds to the system is variable. In a recent visit to Raimat vineyards in Spain (http://raimat.com/es/), the manager of 1800 cultivated lands indicated that they are the owner of 6 harvest machines. Still, in the moment of harvesting, they also rent three more. Asked about the necessary number of devices to finish the work according to their production planning, he indicated that three units are enough, but they need the other options available because the cost of not having them may be very high.

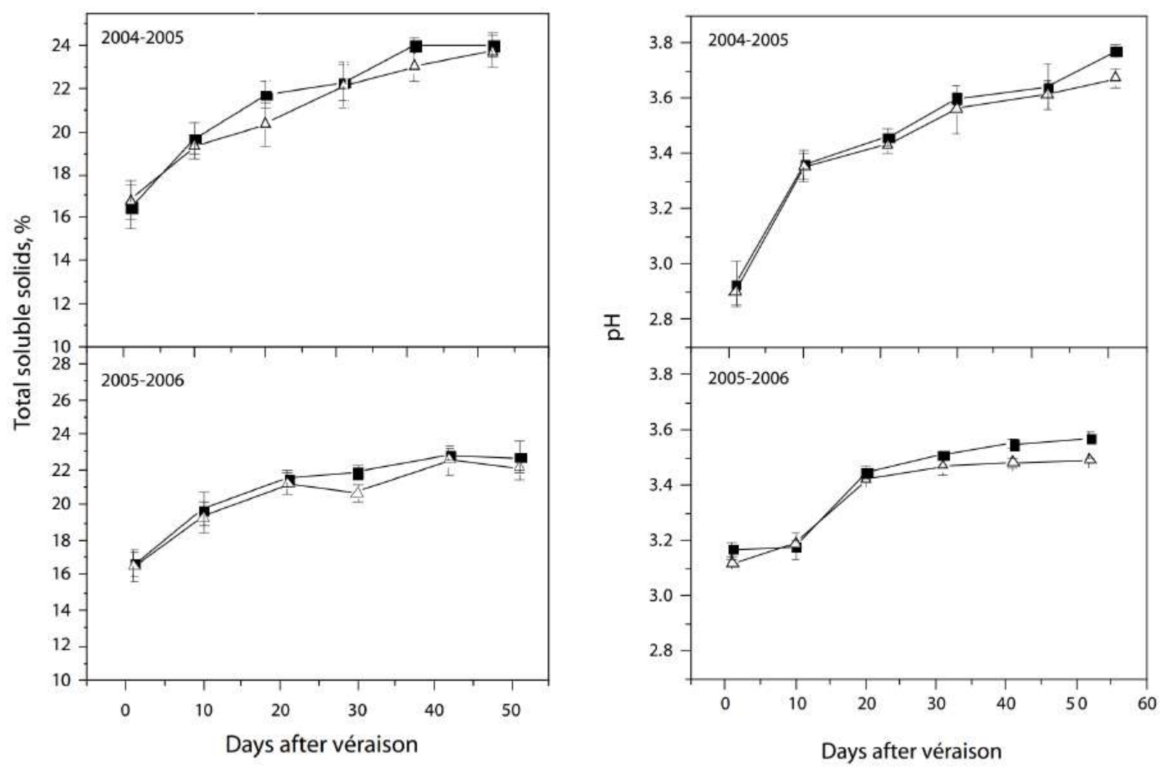
For large extensions, mechanized harvesting could be a reasonable strategy, even more, when the machine's capacity is equivalent to 40 to 50 hand pickers and the cost is less than 20% of the equal manual labor. In the case of little plantations, there is the opportunity to rent machines in a cooperative way to take advantage of cost, avoiding the financial cost of the immobilized money of buying the equipment (usd 300,000 - 600,000). In this case, the schedule of the work is a deal because the postponement or anticipation of the harvest impacts the individual finances, generating a loss of quality in the decision because there is no common good to pursue.

Another essential zone of wine grapes production is Huesca Region. Some producers indicate that their machines work in a mixed regime, both in their fields and in rental for third parties. They reduced the quality of the grape plants to extract with cheaper technology avoiding the cost of the high-quality equipment, making profitable also the rental for other minor farmers of the region. In both cases, despite the presence of machinery, a small portion, around 5-10%, is made by manual labor, because of the conditions of the land (topography).

In Chile, the land extension is smaller than in Spain, and the topography is more irregular. Chile produces different types of wine because of the altitudes of the plantations and their latitudes. According to the local industry information, the harvest process is manual at least 60% of the total product. It is a practical way to manage little farms because it gives the possibility of negotiating the labor cost, reducing fixed costs. As grape needs permanent dedication during the year and not only in the harvest process, a part of the workforce is permanently employed. The required labor for the harvest is a function of the progress of the ripening process, the forecasting capacity, winery space, and, finally, workforce availability. The harvest could last three months, but the actual demand should be balanced with other factors, making difficult the consistency among decisions. To face this, a decision model that supports this process could be advantageous.

**Quality and Ripening process**

The appearance of the germination capacity gives physiological maturity; For industrial areas, the selected maturity is technological or related to functionality in winemaking.
The quality of the *Vinis vinifera* is closely linked to the production area (climate and soil), effective average temperature, water regime, rainfall during or close to harvest, and topography of the land. The volatile organic compounds play a critical role in the final quality of the wine; however, the balance among the environment, vineyard practices, and genotypes is poorly understood (Lund and Bohlmann, 2006). This balance has to be respected when the harvest time is chosen, considering acidity, sweetness, taste, and phenolic ripeness. The quality of the grape at the moment of harvest is the main factor for the quality of the wine (Coombe, 1992). It isn't easy to define the optimal maturity because there is not a privileged element in the chemical profile to set it Meléndez et al. (2013). For example, sugar content and acidity are two very used, but the ratio between them changes depending on the type of wine and the winemaker objective. At least two types of maturity are recognized: technological and phenolic. The first, technological maturity, occurs when the sugar content is very high, so the maturity index (sugar/acidity) is high. This indicator is usual in industry contracts for standard grapes; acidity is measured by $pH$ (with scale




Brix evolution $\quad\quad$ $pH$ evolution

Figure A.5. Brix and $pH$ evolution over time since the *veraison*.
Source:Falcão et al. (2008)

$[1, 14]$, where $pH = 1.0$ is extremely acid, $pH = 6.0$ is neutral, and $pH = 14.0$, extremely alkaline) and sugar contents by $^{\circ}Brix$. The sugar content is transformed into a probable alcoholic equivalent degree considering 16.83 g of sugar/l per $1\%$ alcohol (Glories et al., 2000), and hence its importance. Figure A.5 is shown an example of their progress on time.

The phenolic maturity is reached at the stage when both anthocyanin compound concentration is maximum, and tannins content is low in skin and seed (Le Moigne et al., 2008). Another indicator of maturity is the status of the skin of the grain; following the case of other fruits, the skin becomes softer according to the ripening process. Beyond these "well-documented" ways of ripening evaluation, in practical terms is necessary to consider a mix of them, especially in products that are designed for higher prices than the standard (Coombe, 1992).

The beginning of the ripening stage is the *veraison*, the moment when the color of the grape grain presents the first change. This moment varies among years, and it's responsible for a significant part of the change in the final harvest day. The ripening process is developed in a particular way in each berry, so the uniformity is not assured. Chemical analyses are developed over the juice of the grains to control the progress of the stage. The sugar content increases until a plateau is reached; the changes after this milestone are explained by the loss of water or the gain of water, but not sugar content changes (Coombe, 1992). Figure A.6 shows a very general scheme of the ripening process through two indicators, as berry size and solid contents, measured as $^\circ Brix$.

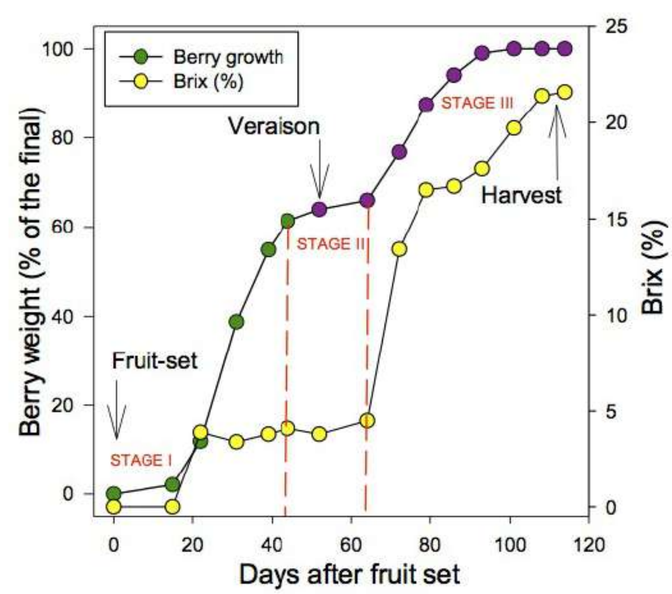



Figure A.6. Berry weight and sugar concentration through different stages. (Source: https://ohioline.osu.edu/factsheet/HYG-1434-11)

The maturation stage brings critical events in the development of this raw material. We detail some below:

- Increase in size and weight
- Increase in sugar content, usually to levels close to 200 gr / liter of grape juice
- Decrease in the concentration of acid content
- Color changes towards the specific pigmentation of the vine type

- Increment in the concentration of aromatic and taste substances, responsible for the organoleptic profile. The observation of the maturation process through these parameters is called phenolic maturation.

According to (Ferrer et al., 2008), the harvest should be done at the right moment; otherwise, value is destroyed. Premium wine quality is degraded if the harvest moment differs from the optimal day. In that paper, the authors represent this behavior with a loss quality function, based on the Taguichi model (see Besterfield (2003)), introducing a novel model. The proposed curve looked alike a parabola and was made using enologist surveys which described the changes following their professional standpoint. In the vinification process, sugar becomes alcohol in a controlled fermentation. Both the alcohol content and high acidity (low pH) allow the wine to be preserved from unwanted microbiological events. Still, they are also part of the desired character of the product.

**Industrialisation**

The industrialization of the vinifera grape is called *vinification*. According to the type of strain to be processed, there are some differences. In figure A.7, (Jackson, 2008) shows the operational flow in the industry for both strains.

The general process includes an initial destemming, common to all types of grapes, removing leaves, stones, and any extraneous material. Immediately after, the crushing process gives juices and waste. Press machines could make the crushing step. The most common is the rotary press, where two objectives are pursued. The extraction of juices is through centrifugal forces and mechanical forces (grains crash with the device's walls). The filtration of the juice is helped by the accumulation of grape waste that stays adhered to the metallic wall because of the centrifugal force. This cake improves the filtering performance, but according to time goes by, it is necessary to remove them to prevent the internal pressure from increasing and losing the working capacity. After this process, each strain presents different approaches in the maceration step. The maceration step facilitates the extraction of minor components from the solid parts of the grape. It's a bioprocess that leads to the must or grape macerate.

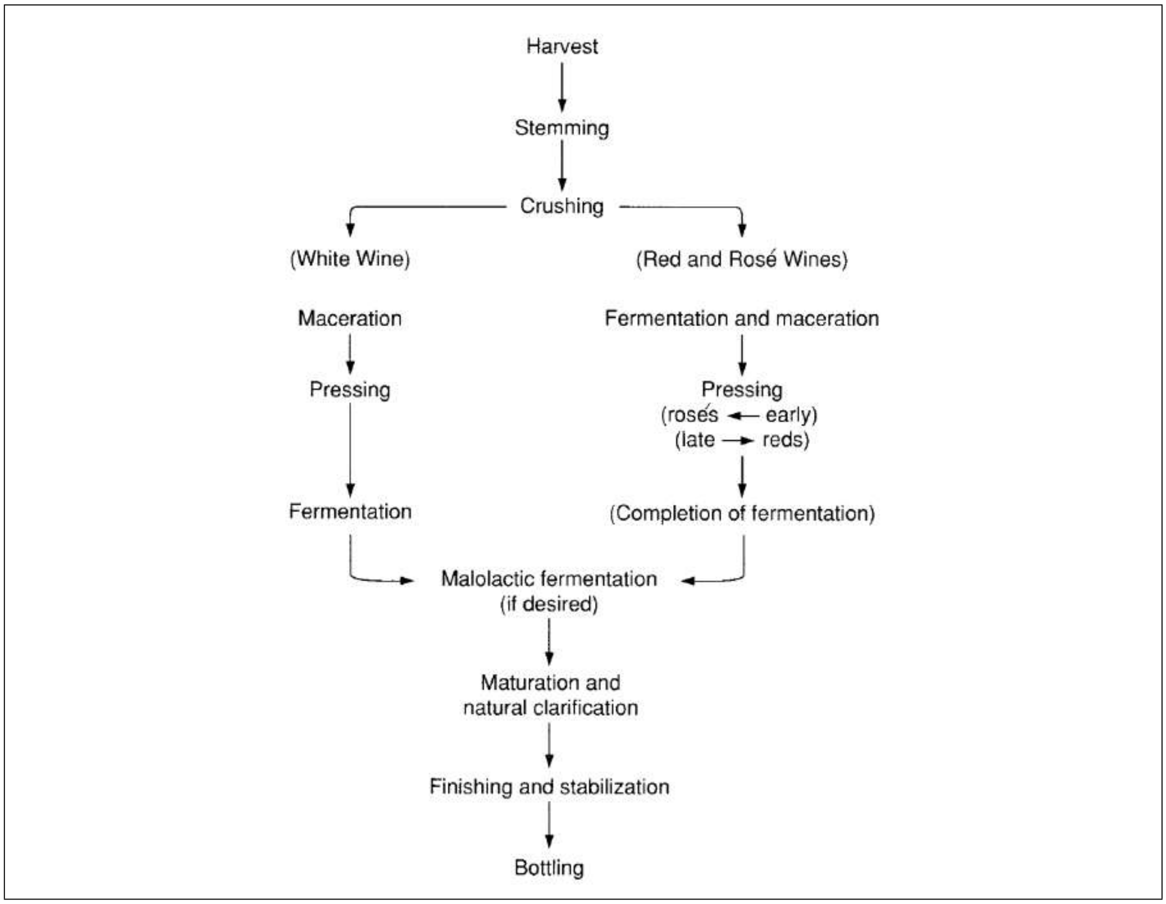



Figure A.7. Diagram Flow for industrialization of both strains. Source: (Jackson, 2008)

For **white strains**, the maceration time lasts only a few hours, and it's kept to a minimum, leaving the primary transforming process to the fermentation stage. The crushed wastes must be processed, so there are three different approaches to recovering the grape must.

- From the crushing machine to the press
- Preprocess with temperature reduction, decreasing the fermentation speed (it is directly proportional to the temperature), and then send to the press
- Start a maceration to achieve better organoleptic characteristics and then continue pressing.

Variants two and three represent extra actions, making the process more expensive than the first approach. The third option is costly, making it a practice for high-quality wines.

In the case of the **red strains**, the maceration and fermentation happen at the same time. The fermentation process increases the ratio of ethanol, and it serves as a solvent in the extraction of phenolic and minor components. The process could extend for several hours. In this stage, the sugars are converted into ethyl alcohol and carbonic anhydride, generating a bubbling from the latter's escape. The microbiological action is key in this stage, with a base, temperature, and density. The temperature is a parameter that allows the biological continuity of the yeasts, and the density control refers to the periodic verification of the existence of sugars, fundamental in the fermentation process. When the sugar level is low, the vat is technically exhausted. The tank used can be made of different materials, which generates other profiles in the final product. In the case of white vines, the time spent in barrels after fermentation can last up to four months.

The free-run juices flow aways under gravity, and the solid material is sent to the pressing step, where the remaining liquid and other components are extracted. The final blends among the different press outcomes and original juice depend on wine type. The must from the press can be tempered in a heat exchanger to cool it down. This must require decantation to separate impurities, in addition to allowing aeration. It is critical to consider the atmosphere for this decantation due to the possible degradation effects of the final product. Post decantation, the solids that have remained at the tank's bottom are filtered to recover the grape juice.

The fermentation process in the red strain is similar to the white strain, but it differs in the working temperature, $20 - 36°$ for red strains. In addition, the must is present in this process and is recirculated periodically.

After completing the alcoholic fermentation, the malolactic fermentation begins. This fermentation consists of the transformation of the malic acid into lactic acid. Lactic acid is a weak acid, so the medium becomes more alkaline. This condition is beneficial for red strains, and grapes from cool regions. White wines take advantage of higher acidity, so this fermentation should be rigorously analyzed in order to be implemented. There are alternative practices, as early clarification, that could lead to competitive results.

The last process of the chemical transformation is the addition of a preservative, in this case, sulfur anhydride, which aims to limit microbial activities in the already prepared wine.

Before proceeding to bottle the wine, there are stages of completion. Clarification is the first of them and tries to eliminate substances that may negatively affect the aspect and taste of the product. Fine elements are separated; they transfer both cloudy appearance and potential off-flavor and odors. The applied technologies differ, using activated carbon to flocculating features that allow the fines' accumulation and subsequent removal. To complement this activity, a traditional filtration or others that improve the relative acceleration of the particles, such as centrifuges to achieve the separation of *finos*, may be used. It is worth mentioning that the latter incorporates air and can lead to accelerated oxidation. The next stage is stabilization, which will allow the wine to extend its shelf life, especially when faced with temperature changes.

Once the wine has finalized its final adaptation, it is subjected to the fractionation or bottling process. The traditional sequence begins with washing the bottles, then the filling that may require a modified atmosphere to reduce the amount of oxygen present, and then the corking, which can be made of natural or synthetic cork. In modern lines, a decoupling of the last stages or finishing of the bottles is generated, consisting of encapsulating the bottle's neck, labeling, boxing, and palletizing.

The wine process has been very well studied. It's only a very briefly introduction, where general features have been indicated for the non especialist reader. In order to go deeper, we recomend Jackson (2008), Boulton et al. (2013), Reynolds (2010), Morata (2018) or Glories et al. (2000).

## B. ABOUT NESTED DISTANCE

### Nomenclature

$\hat{\xi}_t$, means the value of the uncertain event realization at time $t$

$\mathcal{P}_i$,probability spaces characterized by the $i-th$ probability distribution

$\mathbb{T}_i$, tree that represents the $i-th$ probability space

$\mathbb{T}_{ref}$, tree that represents the reference for the planned experiences

$\mathbb{T}_{exp}$, tree that represents the experience

$s \in \mathcal{S}$, means one stage of the set of stages, $St$

$pred(n)$, $n_s^*$ direct predecessor of node $n$ in a tree representation

$pred_s(n)$, predecessor of node $n$ in a tree representation in the stage $s$

$\mathcal{N}$, the complete set of nodes of a tree

$\mathcal{N}_t$, means the list of nodes in at time $t$

$n_s^* \prec n$, means that there is a time $t$ that is a stage $s$ where $n_s^* = pred(pred(...(pred(n))...))$

$\mathring{n}$, denotes a node that belongs to the set $\mathcal{N}_T$

$\mathfrak{F}_t$, filtration in a time $t$

$\xi_n$, the uncertainty event value at node $n$

$\boldsymbol{d}(i,j)$, the value of the nested distance algorithm where $i-th$ is the id number of the first tree to compare, and $j-th$ the id number of the second.

$\hat{\boldsymbol{d}}(i,j)$, the value of the Wasserstein distance where $i-th$ is the id number of the first tree to compare, and $j-th$ the id number of the second.

$\theta_\beta$, Expert/rookie nominal productivity ratio

$\theta_c$, Expert/rookie cost ratio

$\mathbb{M}$, total expected manpower requirement

$\theta_m$, expected percentage of experts manpower

The *nested distance* concept, ND, was introduced by Pflug (2010). In this section, we will introduce the concept (please refer to Pflug and Pichler (2016) for more details).

Nested distance is a multistage generalization of the Wasserstein distance, WD, denoted by $\dot{\boldsymbol{d}}$. Having two probabilities space, $\mathcal{P}_1$ and $\mathcal{P}_2$, there is a cost function $c$ to transport $\mathcal{P}_1$ to $\mathcal{P}_2$. In practice, both are discrete (even when they could be continuous originally), so the transport is between two sets of points. The cost function represents the effort to transport one distribution to other, and it's connected with the value of the samples and their probabilities. If both distributions are equal, the transport will be null, so WD is null. According distributions become different both in probability and values, the WD value increases. The WD is a cost function, where the distance is proportional to the cost. Distances $\dot{\boldsymbol{d}}(\mathcal{P}_1, \mathcal{P}_2$ and $\dot{\boldsymbol{d}}(\mathcal{P}_2, \mathcal{P}_1$ are similar, one of the requirement that a metric should preserve; WD meets all the metric features.

In a multistage tree, leaves $\omega$ are paths where the final state forms a distribution, using conditional probabilities. If we applied WD in a multistage tree, the structure of information and the time evolution are not taken into account, so even when the final state could be similar, the history is lost. WD works like a picture, more than as a movie, in terms of the captured information; used recursively, it should keep the dynamic of the distribution change over time. This is the idea behind ND, an extension of WD to stochastic processes. Now, it's necessary to establish a few concepts. Our goal is not to treat ND mathematically but to generate a strong enough intuition for the application to be correct.

Pflug and Pichler (2016) defines a tree as a representation of a multistage stochastic optimization program in a finite probability space. The probability space for the $i^{th}$ distribution is denoted by $\mathcal{P}_i$, and the tree representation of the stochastic process is $\mathbb{T}_i$. A tree is a directed graph with a single root, where each node belongs to a specific moment in time and contents a set of values for different elements of interest in the problem. Its topology is important because it summarizes how information and decisions are connected. The arcs between nodes have different weights or probability of happening. Consider that a tree has $\mathcal{N}$ nodes, where $n = 1$ is the root. For each node $n > 1$, there is a direct predecessor node, $pred(n)$, or shorter $n^*$. The distance from node $n$ to $n^*$ is called stage (the stage is in time units). Being $\mathcal{N}$ the set of nodes of a tree, $\mathcal{N}_t$ means the list of nodes in that period. In the beginning, $\mathcal{N}_0 = \{1\}$, the root; $\mathcal{N}_T$ means the leaves (final) nodes, and $\mathcal{N}_{0:T-1}$ are the inner nodes or complements of the leaves nodes. If $s$ is a stage of $S$, and we are in a

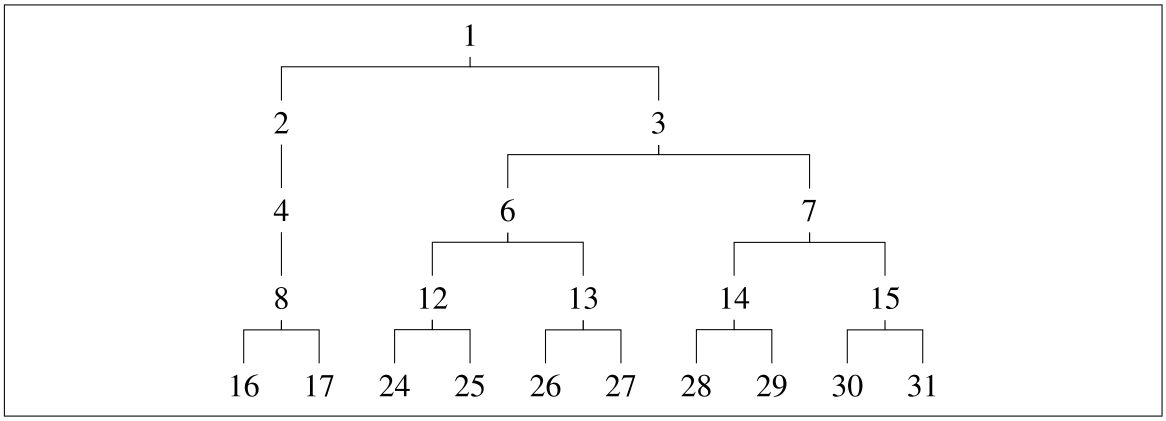



Figure B.1. Tree with unbalance bushiness

period $t$, where $s = t + 1$, then a node $n \in \mathcal{N}_t$, has an immediate predecessor $pred_s(n)$ in period $s$, $n_s^*$, where $n_s^* \in \mathcal{N}_s$ . Considering that the distance between $s$ and $t$ may be more than 1 period, $n_s^*$ is the nested predecessor according to the structure of the tree, $n_s^* = pred(pred(...(pred(n))...))$, where $n \in \mathcal{N}_t$. This formulation allows to consider periods in which there are no decisions, but implementations and thus, new status. A general way of representing this is to write $n_s^* \prec n$ that means that there is a time $t$ where $n_s^*$ precedes $n$. Additionally, it gives the opportunity to describe a path between the decision period and specific status of the system at time. In figure B.1 we can see an example. For node $n : 8$, $pred(8) : \{4, 2\}$; for $n : 16$, $pred(16) : \{8\}$. Stages refer to decision times. If in every period of time, a decision is made, the stages are similar to the periods of time, in number and sequence. But in a more general way, they are not necessary equal. In the figure case, node $n : 8$ is not a direct predecessor of node 8, because between $n : 1$ and $n : 8$ exist alternative path the sub-tree that begins in node $n : 3$, product of a stage or decision moment.

Each node has an unconditional probability of occurrence, while arcs probability is conditional. If $n \in \mathcal{N}_T$, we decorate the nodes as $\mathring{n}$; the conditional arc probability is defined as the probability of being $n$ constrained to the previous path. If the path between two nodes is unique, then the conditional probability is referred to have been in a specific predecessor, $n_s^*$. The conditional probability $w(n) = P(n \mid n_s^*) = P(n)/P(n_s^*)$.

If we consider the tree process (based on a stochastic process), appears an equivalence between tree structure and filtration concept $\mathfrak{F}$. Filtration is linked to the information that

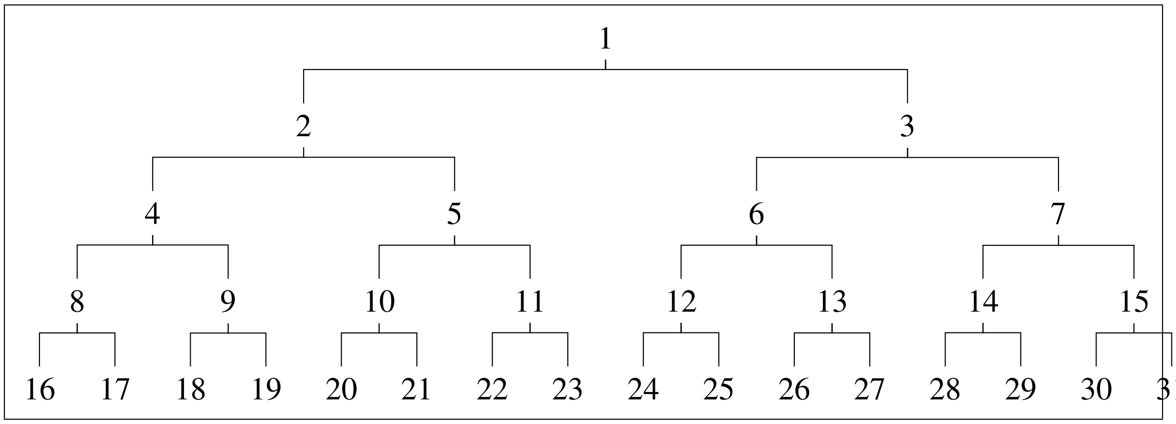



Figure B.2. Binomial Tree structure Example, with 16 leaves and 4 stages

is available at a specific time. As time goes by, the filtration is bigger; if $\mathfrak{F}_t$ means the filtration at moment $t$, then $\mathfrak{F}_1 \leq, ..., \leq \mathfrak{F}_{t-1} \leq \mathfrak{F}_t$; at a specific moment, $\mathfrak{F}_t$, there is a natural filtration that considers the whole past information.

For making the concept clearer, see figure B.2, a binomial tree structure. The same tree is written like a table in table B.1. Each leaf corresponds to a node in $\mathcal{N}_T$, with $T = 4$. For example, the path to the $n = 30$ or leaf $\omega_{15}$ is $(1, 3, 7, 15)$, the collection of nodes that precedes the final node. We have four stages, where the set of nodes are $\mathcal{N}_1$, $\mathcal{N}_2$, $\mathcal{N}_3$ and $\mathcal{N}_4$.

In terms of filtration, the tree is $\mathfrak{F} = \{\mathcal{F}_0, \mathcal{F}_1, \mathcal{F}_2, \mathcal{F}_3\}$, that is the set of information available at each moment, and the definition of each of these $\sigma$-algebras are:

$\mathcal{F}_0 = \sigma(\{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6, \omega_7, \omega_8, \omega_9, \omega_{10}, \omega_{11}, \omega_{12}, \omega_{13}, \omega_{14}, \omega_{15}, \omega_{16}\})$

$\mathcal{F}_1 = \sigma(\{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6, \omega_7, \omega_8\}, \{\omega_9, \omega_{10}, \omega_{11}, \omega_{12}, \omega_{13}, \omega_{14}, \omega_{15}, \omega_{16}\})$

$\mathcal{F}_2 = \sigma(\{\omega_1, \omega_2, \omega_3, \omega_4\}, \{\omega_5, \omega_6, \omega_7, \omega_8\}, \{\omega_9, \omega_{10}, \omega_{11}, \omega_{12}\}, \{\omega_{13}, \omega_{14}, \omega_{15}, \omega_{16}\})$

$\mathcal{F}_3 = \sigma(\{\omega_1, \omega_2\}, \{\omega_3, \omega_4\}, \{\omega_5, \omega_6\}, \{\omega_7, \omega_8\}, \{\omega_9, \omega_{10}\}, \{\omega_{11}, \omega_{12}\}, \{\omega_{13}, \omega_{14}\}, \{\omega_{15}, \omega_{16}\})$

$\mathcal{F}_4 = \sigma(\{\omega_1\}, \{\omega_2\}, \{\omega_3\}, \{\omega_4\}, \{\omega_5\}, \{\omega_6\}, \{\omega_7\}, \{\omega_8\}, \{\omega_9\}, \{\omega_{10}\}, \{\omega_{11}\}, \{\omega_{12}\}, \{\omega_{13}\},$
$\{\omega_{14}\}, \{\omega_{15}\}, \{\omega_{16}\})$

The filtration at the beginning considers that any of the scenarios are possible. According to time goes by, the possible scenarios are gathered by the nodes because the available information increases each period of time. The filtration is connected with the structure

Table B.1. Tree process written as a table

| Leaves | $t_0$ | $t_1$ | $t_2$ | $t_3$ | $t_4$ |
|---|---|---|---|---|---|
| $\omega_1$ | 1 | 2 | 4 | 8 | 16 |
| $\omega_2$ | 1 | 2 | 4 | 8 | 17 |
| $\omega_3$ | 1 | 2 | 4 | 9 | 18 |
| $\omega_4$ | 1 | 2 | 4 | 9 | 19 |
| $\omega_5$ | 1 | 2 | 5 | 10 | 20 |
| $\omega_6$ | 1 | 2 | 5 | 10 | 21 |
| $\omega_7$ | 1 | 2 | 5 | 11 | 22 |
| $\omega_8$ | 1 | 2 | 5 | 11 | 23 |
| $\omega_9$ | 1 | 3 | 6 | 12 | 24 |
| $\omega_{10}$ | 1 | 3 | 6 | 12 | 25 |
| $\omega_{11}$ | 1 | 3 | 6 | 13 | 26 |
| $\omega_{12}$ | 1 | 3 | 6 | 13 | 27 |
| $\omega_{13}$ | 1 | 3 | 7 | 14 | 28 |
| $\omega_{14}$ | 1 | 3 | 7 | 14 | 29 |
| $\omega_{15}$ | 1 | 3 | 7 | 15 | 30 |
| $\omega_{16}$ | 1 | 3 | 7 | 15 | 31 |

of the tree, considering the possible scenarios or leaves. A different structure gives other filtrations.

In terms of the original work by Pflug and Pichler (2016), the structure $(\Omega, \mathfrak{F}, P, \xi)$ is called a value-and-information structure. The adding $\xi$ represents the value that the structure takes in each node, introducing the node-orientated notation, $\xi_n$ for $n \in \mathcal{N}$.

In this work, we mapped the tree space exhaustively with the complete node collection. As Timonina (2015) mentions, "...the tree that represents the finitely valued stochastic process, is called finitely valued tree. To solve the approximate problem numerically we should represent the stochastic process $\xi$ as a finitely valued tree. ". In the same way, we use the information of the decisions policies to complete the knowledge of the tree and then proceed with the ND algorithm.

To estimate the ND we follow the algorithm proposed in Pflug and Pichler (2016)

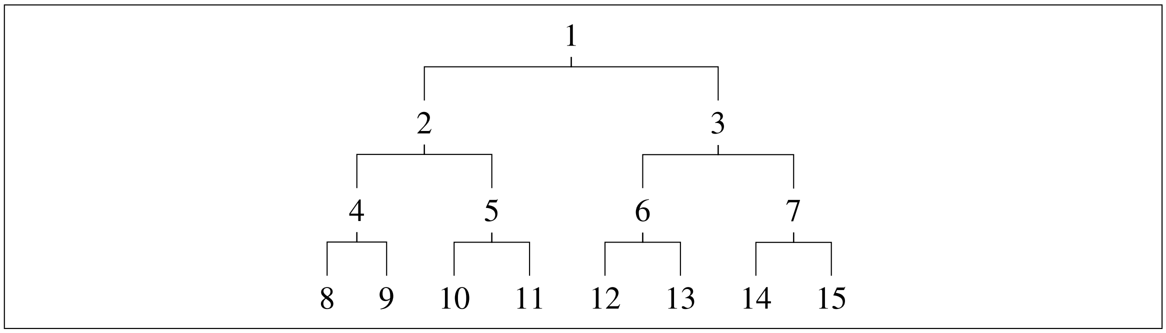



Figure B.3. Binomial Tree structure Example, with 8 leaves and 3 stages

According to it, we calculate in a backward iteration from the last stage to the first. In our case, the last stage is $T$ because we include in each period the final status of the system. Hence, the first stage is $t = 1$. The period of interest (estimate how different both trees are) is $[T, ..., 1]$. ND has been used to represent the difference in stochastic processes as a measure of distance. Once that a stochastic process is mapped completely (a developed tree), ND could be implemented.

The ND value is a scalar that represents a metric to characterize the relative distance. We call this scalar the value nested distance for both specifics trees, and the notation is $\boldsymbol{d}(i, j)$ where $i^{th}$ and $j^{th}$ are the id number trees. For example, if $\boldsymbol{d}(1, 2) = 100$ and $\boldsymbol{d}(1, 3) = 1000$, we can say that trees $\mathbb{T}_1$ and $\mathbb{T}_2$ are more similar than $\mathbb{T}_1$ and $\mathbb{T}_3$. Similar trees, with little changes, could present very different ND; to gain better intuition about the contribution of information and structure in ND value, we explored a few controlled experiences below.

A binomial tree of three stages is shown in figure B.3. In the tree are represented the nodes that are numbered successively. The nested distribution of that tree is shown in figure B.4. Each node is replaced by the unconditional probability of the node and its information, in this case, a variable. In this case, probabilities are similar among branches, and the variable's value is also the same. Only in the initial stage, the probability of the node is one because it is unique. The nested level of the boxes indicates the number of stages. A reference tree similar to the shown in figure B.3 is called $\mathbb{T}_{ref}$. We keep it constant during the three experiences. Now, we define the three experiences that we carried out.

<table>
<tr><td colspan="8" align="center">1</td></tr>
<tr><td colspan="8" align="center">10</td></tr>
<tr><td colspan="4" align="center">0.5</td><td colspan="4" align="center">0.5</td></tr>
<tr><td colspan="4" align="center">10</td><td colspan="4" align="center">10</td></tr>
<tr><td colspan="2" align="center">0.5</td><td colspan="2" align="center">0.5</td><td colspan="2" align="center">0.5</td><td colspan="2" align="center">0.5</td></tr>
<tr><td colspan="2" align="center">10</td><td colspan="2" align="center">10</td><td colspan="2" align="center">10</td><td colspan="2" align="center">10</td></tr>
<tr><td>0.5</td><td>0.5</td><td>0.5</td><td>0.5</td><td>0.5</td><td>0.5</td><td>0.5</td><td>0.5</td></tr>
<tr><td>10</td><td>10</td><td>10</td><td>10</td><td>10</td><td>10</td><td>10</td><td>10</td></tr>
</table>

Figure B.4. Tree Nested distribution Example

- Experiment I: We define tree $\mathbb{T}_{exp}$, similar to $\mathbb{T}_{ref}$ but for $n:2$. The value of that node will be different from $10$; a sensitivity analysis is accomplished by changing the value in the range $[-40\%, +40\%]$, and $\boldsymbol{d}(ref,\, exp)$ is determined. To complete this first experiment, we also change the probability of $n:2$, and consequently $n:3$. We make the sensitivity analysis for the range $[-40\%, +40\%]$.
- Experiment II: We define tree $\mathbb{T}_{exp}$, similar to $\mathbb{T}_{ref}$ but for $n:5$. The value of that node will be different from $10$; a sensitivity analysis is accomplished by changing the value in the range $[-40\%, +40\%]$, and $\boldsymbol{d}(ref,\, exp)$ is determined. To complete this first experiment, we also change the probability of $n:5$, and consequently $n:4$. We make the sensitivity analysis for the range $[-40\%, +40\%]$.
- Experiment III: We define tree $\mathbb{T}_{exp}$, similar to $\mathbb{T}_{ref}$ but for $n:11$. The value of that node will be different from $10$; a sensitivity analysis is accomplished by changing the value in the range $[-40\%, +40\%]$, and $\boldsymbol{d}(ref,\, exp)$ is determined. To complete this first experiment, we also change the probability of $n:11$, and consequently $n:10$. We make the sensitivity analysis for the range $[-40\%, +40\%]$.

In figure B.5 the first experiment are presented. To simplify the notation, we will call $a$ the node that is under study. When $\mathbb{T}_{exp}$ is similar to $\mathbb{T}_{ref}$, the $\boldsymbol{d}(ref,\, exp)$ is zero. If we keep constant the probability in $a$ as in $\mathbb{T}_{ref}$, according $n:2$ relative change increases, also the

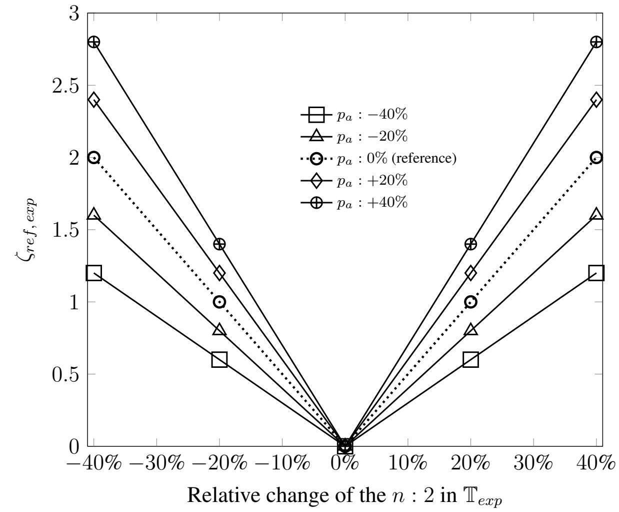
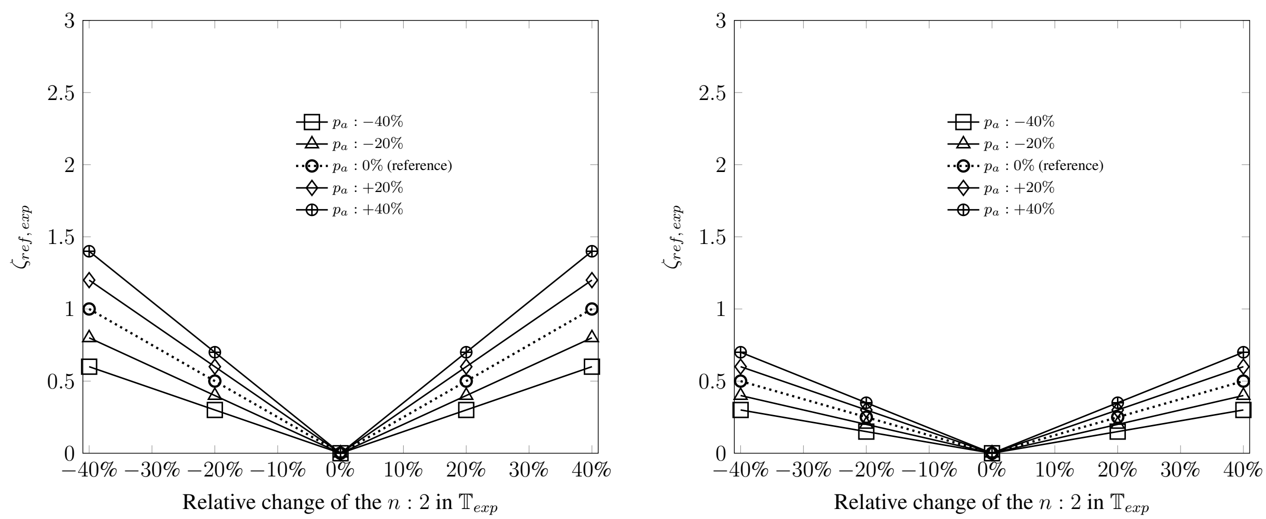



Figure B.5. Nested Distance for the first experiment




(a) Second Experiment

(b) Third Experiment

Figure B.6. Nested Distribution Example

$\boldsymbol{d}(ref,\, exp)$ does. The curve is symmetric because the distance is the same in absolute number. About probability, the impact is a little different. If probability in $a$ increases, the nested distance is more sensitive. This behavior is because the importance of the node is greater in the tree value. If the probability of occurrence of the abnormal (considering the pattern in $\mathbb{T}_{ref}$) is minor, the potential effects are diluted, so the $\boldsymbol{d}(ref,\, exp)$ become shorter.

In figure B.6, the two remaining experiment results are shown. The general behavior appears very likely to first experience. Still, the extension of $\boldsymbol{d}(ref,\, exp)$ is minor since the node is to be disrupted closer to the final periods of the time horizon. It is intuitive because the effect of the change has less time to expand its impact. In this example, we conclude that:

- The nested distance increases according to the difference between them is earlier
- While bigger is the node value difference, the nested distance also.
- Probability impacts more if its relative change benefits the node suffering the alteration in its value.

The experiences insight does not allow to connect the model performances with the nested distance or even the nested distance value with the internal information. Vitali (2018) indicates that the objective function value in a multistage stochastic problem is positively correlated to the nested distance, but there is no information about more concrete relationships. They go even further in Horejšová et al. (2020), and the conclusions are in the same line. We think that there is a space for an exhaustive study of the numeric behavior of nested distance, but it's out of this thesis' scope. Finally, for measure instances, ND has proved valid and meaningful in relatives analysis in the literature.

The implementation of the algorithm for ND was made in Python 3.0, using a python library to calculate the WD, called POT Flamary et al. (2021), where the algorithm approximates the Sinkhorn instead of WD as it's described by Cuturi (2013).